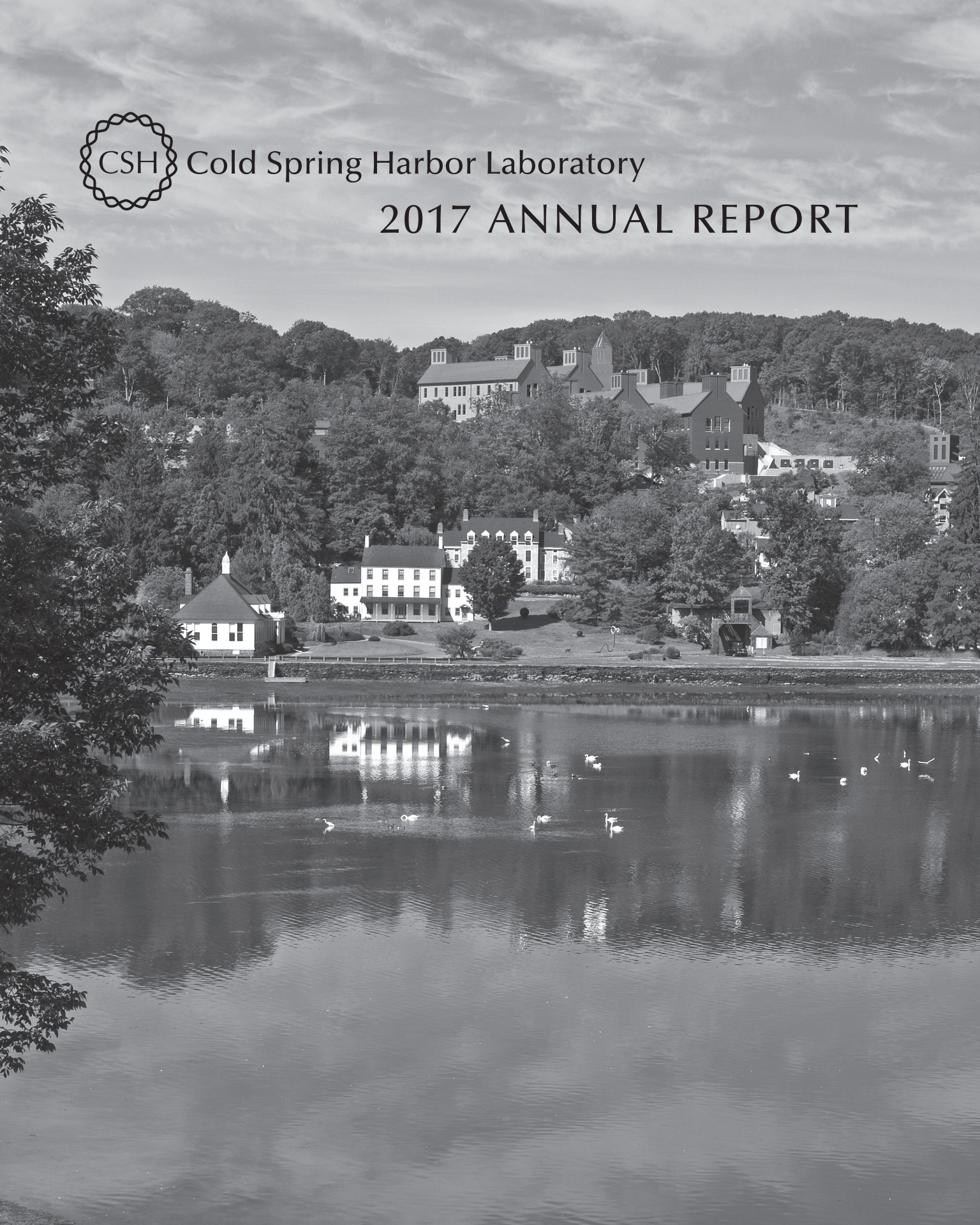

CSH
Cold Spring Harbor Laboratory
2017 ANNUAL REPORT


CSH
Cold Spring Harbor Laboratory
2017 ANNUAL REPORT

ANNUAL REPORT 2017

Cold Spring Harbor Laboratory
One Bungtown Road
Cold Spring Harbor, New York 11724

www.cshl.edu

Managing Editors Philip Renna, Dagnia Zeidlickis
Production Editor Kathleen Bubbeo
Science Writing Peter Tarr
Copy Editors Dorothy Brown, Joanne Dolce
Production Manager Denise Weiss
Nonscientific Photography Constance Brukin, Gina Motisi
Cover Designer Margot Bennett

Front cover: Photography by Constance Brukin

Contents

Officers of the Corporation

Jamie C. Nicholls, *Chair*
Robert D. Lindsay, *Vice-Chair*
Marilyn H. Simons, Ph.D., *Vice-Chair*

Leo A. Guthart, *Treasurer*
Paul J. Taubman, *Secretary*

Bruce Stillman, Ph.D., *President and CEO*
W. Dillaway Ayres, Jr., *Chief Operating Officer*

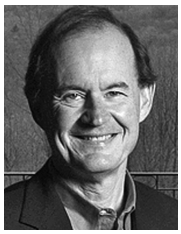
Board of Trustees

Individual Trustees

Lalit R. Bahl, Ph.D.
Senior Research Scientist, Renaissance Technologies Corp.

Tania A. Baker, Ph.D.
E.C. Whitehead Professor, Massachusetts Institute of Technology; Investigator, HHMI

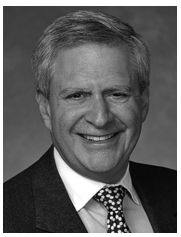
Joanne Berger-Sweeney, Ph.D.
President, Professor in Neuroscience, Trinity College



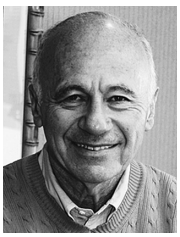
David Boies
Chairman, Boies, Schiller & Flexner LLP

Michael R. Botchan, Ph.D.
Richard and Rhoda Goldman Distinguished Professor of Biochemistry, University of California, Berkeley

Michele Munn Celestino
Laurel Hollow, New York



Charles I. Cogut
Senior M&A Counsel, Simpson Thacher & Bartlett LLP



Leo A. Guthart
Founder and CEO, Topspin Partners

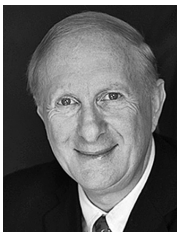
Jeffrey E. Kelter
Senior Partner and CEO, KTR Capital Partners

Laurie J. Landeau, V.M.D.
General Manager, Listowel, Inc.

Robert D. Lindsay
Co-Managing Partner, Lindsay Goldberg

Robert W. Lourie, Ph.D.
Head of Future Research, Renaissance Technologies Corp.

Elizabeth McCaul
Partner-in-Charge, New York Office, Promontory Financial Group, LLC



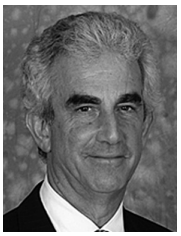
Howard L. Morgan, Ph.D.
Partner, First Round Capital; Director, Idealab

Jamie C. Nicholls
Chairman of the Board, New York, New York

John C. Phelan
Managing Partner, MSD Capital, L.P.

Thomas C. Quick
Palm Beach, Florida

William S. Robertson
Robertson Foundation for Government



George Sard
Chairman and CEO, Sard Verbinnen & Co.

Thomas A. Saunders III
CEO, Ivor & Co., LLC

Charles L. Sawyers, M.D.
Investigator, HHMI; Chair, Human Oncology & Pathogenesis Program, MSKCC

Douglas Schloss
President and CEO, Rexford Management

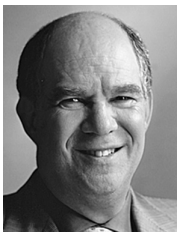
Marilyn H. Simons, Ph.D.
President, The Simons Foundation

Dinakar Singh
Founding Partner, TPG-Axon Capital

Individual Trustees *(continued)*

Samuel L. Stanley, Jr., M.D.
President, Stony Brook University

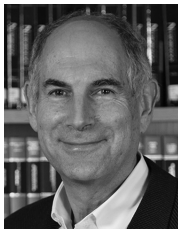
Bruce Stillman, Ph.D.
President & CEO, Cold Spring Harbor Laboratory



James M. Stone, Ph.D.
Chairman, The Plymouth Rock Company

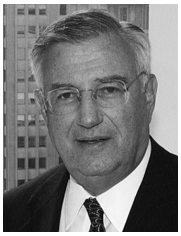
Paul J. Taubman
PJT Partners, LLC

Edward Travaglianti
President, TD Bank Long Island



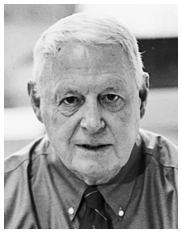
Stuart Weisbrod, Ph.D.
Chief Investment Officer, Iguana Healthcare Partners

George Yancopoulos, M.D., Ph.D.
President, Research Laboratories, & CSO, Regeneron Pharmaceuticals, Inc.

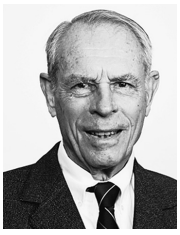


Roy J. Zuckerberg
Senior Director, Goldman Sachs & Co.

Honorary Trustees



Bayard Clarkson, M.D.
Memorial Sloan Kettering Cancer Center



Norris W. Darrell
Senior Counsel, Sullivan & Cromwell LLP

Charles F. Dolan
Cablevision Systems Corp.

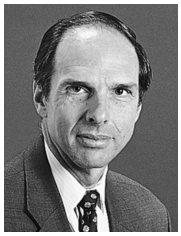
Helen Ann Dolan
Cablevision Systems Corp.

Lola N. Grace
Founder, Middle East Children's Institute

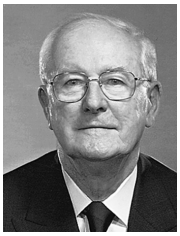
David H. Koch
Koch Industries, Inc.

Mary D. Lindsay
Laurel Hollow, New York

Nancy Abeles Marks
Nancy Marks Interiors and Carl Marks & Co., Inc.



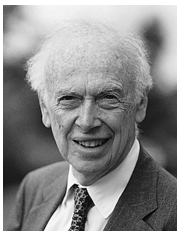
Eduardo G. Mestre
Vice Chairman, Evercore Partners



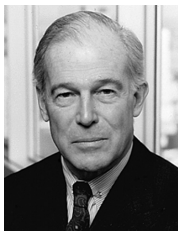
William R. Miller
New York, New York

Alan Seligson
NAK International

James Simons, Ph.D.
Chairman, The Simons Foundation



James D. Watson, Ph.D.
Chancellor Emeritus, Cold Spring Harbor Laboratory



Henry Wendt III
Friday Harbor, Washington

CSHL Scientific Advisory Council

The Scientific Advisory Council (SAC) is an external advisory group that advises the senior management of Cold Spring Harbor Laboratory (CSHL) on matters pertaining to science (both current and future), including the development of a research strategy to maintain CSHL as a world leader. The SAC is a nine-member Council, including a Chair of Council who is an individual known for scientific breadth and a detailed understanding of research management at the senior management level. The other eight members are world leaders in their respective fields and as such are able to provide advice on the different research areas of the Laboratory.

Frederick W. Alt, Ph.D. (Chair)
HHMI Investigator/Harvard University Professor and Co-Chief, Division of Molecular Medicine, Children's Hospital Boston

Angelika Amon, Ph.D.
HHMI Investigator/Massachusetts Institute of Technology Professor, Biology

Joanne Chory, Ph.D.
HHMI Investigator/The Salk Institute for Biological Studies, Professor and Director, Plant Biology Laboratory

Carol W. Greider, Ph.D.
Johns Hopkins School of Medicine, Professor and Director of Molecular Biology and Genetics

Eve Marder, Ph.D.
Brandeis University, Professor of Neuroscience, Biology Department and Volen Center

Markus Meister, Ph.D.
California Institute of Technology, Professor, Molecular and Cellular Biology

Steven Salzberg, Ph.D.
Johns Hopkins University School of Medicine, Professor of Biomedical Engineering, Computer Science, and Biostatistics

Kevan M. Shokat, Ph.D.
HHMI Investigator/University of California, San Francisco; University of California, Berkeley, Professor and Chair, Cellular and Molecular Pharmacology, Professor of Chemistry

Max S. Wicha, M.D.
University of Michigan, Professor and Director, University of Michigan Comprehensive Cancer Center

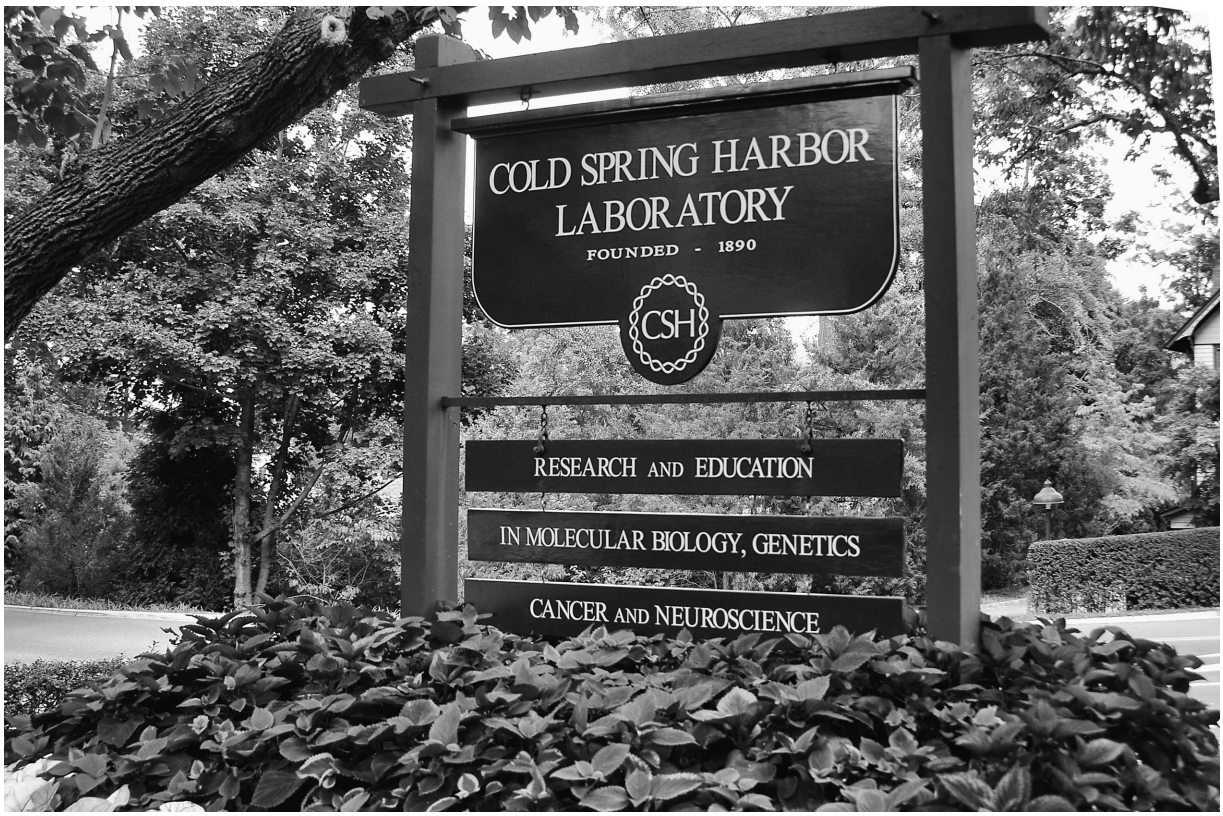
Governance

The Laboratory is governed by a Board of Trustees of up to 35 members that meets three or four times a year. Authority to act for the Board of Trustees between meetings is vested in the Executive Committee of the Board of Trustees. The Executive Committee is composed of the Officers of the Board and any other members who may be elected to the Executive Committee by the Board of Trustees. Additional standing and ad hoc committees are appointed by the Board of Trustees to provide guidance and advice in specific areas of the Laboratory's operations.

Representation on the Board of Trustees itself is divided between business and community leaders and scientists from major educational and research institutions.

The Laboratory is chartered as an educational and research institution by the Board of Regents of the Education Department of the State of New York. It is authorized to operate a graduate program under the name "Cold Spring Harbor Laboratory, Watson School of Biological Sciences" and thereat to confer the degrees of Doctor of Philosophy (Ph.D.), Master of Science (M.S.), and Doctor of Science (Sc.D.), Honorary.

It is designated as a "public charity" under Section 501(c)(3) of the Internal Revenue Code.



Committees of the Board 2017

Executive

Jamie C. Nicholls, Chair
W. Dillaway Ayres, Jr.
Casey Cogut
Leo Guthart
Marilyn H. Simons
Bruce Stillman
James M. Stone
Paul J. Taubman
Edward Travaglianti
Roy J. Zuckerberg

Development

Marilyn H. Simons, Chair
W. Dillaway Ayres, Jr.
David Boies
Casey Cogut
Kristina Davison
Jeff Kelter
Laurie J. Landeau
Howard Morgan
Jamie C. Nicholls
George Sard
Dinakar Singh
Bruce Stillman
Paul J. Taubman
Roy J. Zuckerberg

Commercial Relations

James M. Stone, Chair
W. Dillaway Ayres, Jr.
Casey Cogut
Robert Martienssen
Bruce Stillman

Audit

Leo A. Guthart, Chair
Elizabeth McCaul
Doug Schloss
Edward Travaglianti

Investment

John C. Phelan, Chair
W. Dillaway Ayres, Jr.
Jamie C. Nicholls
Thomas Purcell
William Robertson
Thomas Saunders
Dinakar Singh
Edward Travaglianti
Remy Trafelet

Finance

Leo A. Guthart, Chair
W. Dillaway Ayres, Jr.
Jamie C. Nicholls
Lari Russo
Douglas Schloss
Marilyn H. Simons
Edward Travaglianti

Academic Affairs

James M. Stone, Chair
W. Dillaway Ayres, Jr.
Lalit Bahl
Tania Baker
Michael Botchan
Jacob Goldfield
Laurie J. Landeau
Robert Lourie
Bruce Stillman

Nominating

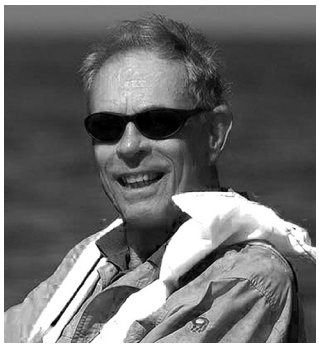
Paul J. Taubman, Chair
W. Dillaway Ayres, Jr.
Casey Cogut
Leo A. Guthart
Robert Lindsay
Elizabeth McCaul
Jamie C. Nicholls
Thomas Saunders
Bruce Stillman
George Yancopoulos
Roy J. Zuckerberg

DNA Learning Center

Laurie J. Landeau, Chair
W. Dillaway Ayres, Jr.
Casey Cogut
Kristina Davison
Jeff Kelter
David Micklos
Jamie C. Nicholls
Ed Rover
Marilyn H. Simons
Bruce Stillman

Robertson Research Fund, Inc.

(Board of Trustees)
W. Dillaway Ayres, Jr.
Walter Goldschmidts
William Gridley
Leo A. Guthart
John Linnartz
Andrew McKibben
Jamie C. Nicholls
William Robertson
Bruce Stillman



Norris W. Darrell (1929–2016)

When remembering Norris Darrell, our neighbor, friend, trustee, and steadfast supporter, I am most grateful for his fervent commitment to advancing higher education. Together with his wife, Henriette, Norris enhanced Cold Spring Harbor Laboratory's dual mission of excellence in scientific research and education, endowing the Watson School of Biological Sciences as well as the Darrell Research Fund to support innovative investigators.

Norris W. Darrell was born in Berlin, Germany, on May 10, 1929. He spent his childhood in New York City, where he was schooled at St. Bernard's. After graduating from Phillips Exeter Academy in New Hampshire, Norris received an A.B. in 1951 from Harvard College. He went on to finish Harvard Law School, cum laude, in 1954, and then served in the U.S. Army until 1956. He returned to civilian life to join the law firm of Sullivan & Crowell, making partner in 1965 and leading the firm's Paris office from 1968 to 1971. Norris achieved much success building the S&C global securities, corporate, and financing law practice.

Along the way, he sought a balance in his life that led him to Austria for ski trips. This is where he met Henriette Maria Haid, to whom he was wed in 1962. In 1965, Norris decided to retire to spend time with his family, which now included a son, Andrew. Another goal of his retirement was to dedicate time to civic organizations, and thankfully to CSHL.

CSHL got to know Norris and Henriette in the 1970s. Norris served on the CSHL Board of Trustees from 1974 to 1981, with terms on the Tenure & Appointments (1976) and Building committees (1976, 1979). This was an exciting time at the Laboratory as the number of research labs proliferated, areas of research expanded to include mammalian genetics, and the cancer and plant science programs grew. This was also a time of expansion for our education programs: New CSHL Press forays into books, lab manuals, and journals met with great success; the Banbury Conference Center was established; and the Meetings & Courses Program grew to attract scientists year-round.

After his board term, Norris was welcomed back in 1986 as a CSHL Associate by the Long Island Biological Association (LIBA). Since 1923, LIBA had raised private funds to support the Laboratory, and today its members are called the CSHL Association. I found a note that Norris penned to fellow CSHL supporters Mary Lindsay and Bill Miller in 1987 that so clearly demonstrated his growing appreciation for the Laboratory: "You and Bill were most eloquent in reminding us of how precious the Lab and strong its claim." Norris wrote that his contributions to the Laboratory at the time, although merely "a token as compared to the major sums needed … [were] intended to demonstrate whole-hearted support, and will come in the belief that it could not be better spent."

But I do believe that I saw Norris most excited when he got involved in our post–high school education initiatives. This was where his heart was, both at CSHL, in his other island home on Martha's Vineyard, and also nationally in his leadership role at the Lumina Foundation. For Norris, providing young people with access to post–high school education was "so necessary today for personal growth and success" (Burrell C, June 6, 2005, *Vineyard Gazette,* "Fund assists students with hardship").

In 1993, he and Henriette were the first to endow an Undergraduate Research Program scholarship, helping fund participation of college students in CSHL's increasingly popular 10-week summer research immersion program. Norris chose to call this the Shakespeare Fund, explaining that he wanted to "suggest the level of achievement expected, as well as the merger of all disciplines at that level." In thanking the Darrells for their philanthropy, then development director Claire Fairman responded, "Shakespeare now has a secure place in science as well as literature."

Norris was part of the Capital Campaign Committee in 1999, helping to raise money to establish CSHL's Ph.D.-granting program—the Watson School of Biological Sciences. In 2002, he and Henriette were proud to endow in perpetuity one of the school's most unique courses, the Darrell Core Course on Scientific Exposition and Ethics. In a 2001 letter, he wrote that "both topics have been of special interest in my professional life."

Norris' enthusiasm for the Laboratory seemed never-ending. And I'd like to think that CSHL held as special a place in his heart as his Atlantic (A 56, Tara) that he raced out of the Cold Spring Harbor Beach Club, where he was an active sailor and tennis player. In 2006, he and Henriette established the Darrell Research Fund as an endowment to support scientific research. For their giving to both our education and research programs, we are ever grateful to Norris and Henriette, who clearly understood how an endowment helps to sustain an institution. Thank you, Norris, for your thoughtful appreciation of how science can be propelled by nimble and flexible investments in new directions. We also thank you for sharing your family with CSHL all these years, as our neighbors across the harbor. They are now forever part of our family.

Bruce Stillman

PRESIDENT'S REPORT

Science has been one of the *most* important contributors to American national strength over the past century, but particularly since the Second World War. During that extraordinary crisis, outstanding national leaders recognized the untapped power of discoveries in a broad range of disciplines—from chemistry and physics to biology and engineering. In 1944, President Franklin Roosevelt commissioned a report from Vannevar Bush, then head of the Carnegie Institution for Science and director of the White House Office of Scientific Research and Development. Bush delivered in 1945 to the then-new President Truman a remarkable document and showed clearly that the country's intellectual capital could contribute powerfully to the success of the war effort and beyond. This crucial national experience provided the rationale for robust postwar federal support for basic scientific research—one of the universally acknowledged sources of modern American economic strength and competitiveness.

The United States has agencies like the National Institutes of Health and the National Science Foundation that are the envy of like agencies in other countries. But in recent years, I have noted various threats to the nation's leadership position in the sciences and the ability of these agencies to operate effectively. These threats include across-the-board federal austerity; Congressional failure to agree upon annual budgets, leading to appropriation of funds far too late in the budget year for sensible planning; and the retreat by industry from investing in difficult areas of R&D like drug development. Over the last year or so, new threats have emerged that are potentially more consequential than any of these. One is the complete lack of appreciation for scientific research in policy decision-making by the current White House administration. Another is changes in immigration policy and the impact such changes might have upon our historic ability to attract the best and brightest young people in the world to America's scientific enterprise, irrespective of nationality.

Attracting the Best

A key source of American greatness today is its long tradition of inclusiveness: the nation's historic ethos of *e pluribus unum*—out of many, one. As in so many other aspects of national life, immigrants have contributed mightily to the leading position of U.S. science. One indicator is the percentage of Nobel Prizes in the sciences won by individuals who lived and worked in the United States at the time of their award, but who were born in other lands. Since the inception of the Prizes in 1901, 95 of America's 289 science laureates (Chemistry, Physics, Physiology or Medicine) have been immigrants—one in three. At Cold Spring Harbor Laboratory we are proud of our association with three of these: Max Delbrück and Salvador Luria, refugees of European fascism, and Richard Roberts, who hailed from England. In 2016, all six of the American science laureates were immigrants, and in 2017, two of seven. Adil Najam, Dean of the School of Global Studies at Boston University, has noted that no single nation can claim as many Nobel Prizes as the group of U.S. immigrant laureates that these eight now join, whose number is exceeded only by the group of Nobelists born in the United States.

We are not only "a nation of immigrants," as the familiar phrase reminds us—in today's world, our dynamism directly reflects our openness. America's leadership in the sciences and its competitiveness as a nation rely on our continuing to be a place where the most talented people want to live, study, and work.

The National Science Board, which annually generates "Science and Engineering Indicators" on behalf of the National Science Foundation, states clearly in its 2017 report what is at stake. "The global landscape of science and engineering research, education and business activities has

undergone dramatic shifts since the turn of the 21st century, as regions, countries and economies around the globe continue to invest in science and technology." Until recently, science and engineering capabilities have been located mainly in the United States, Western Europe, and Japan. In the last 10 to 15 years, they have spread to the developing world—notably to China. Asian nations in particular have invested heavily and at an accelerating pace to build their science and technology capabilities.

As they have made these investments, the global trend has been toward knowledge-intensive activities, including those we engage in at Cold Spring Harbor Laboratory. The economies that lead today's world rely increasingly on a highly skilled workforce and sustained investment in R&D to produce new knowledge, new technologies, and discoveries. In this context, the long-term danger to our leadership in science posed by an administration that shuns science is plain. The danger posed by turning away from other countries or from our long-standing tradition of seeking to attract the world's most talented people should be equally obvious.

Higher education, R&D, and economic activity are closely linked in knowledge economies. Students and researchers are mobile to an unprecedented degree, as are trade and investment. Supply chains and infrastructure are planned and deployed globally. Collaboration and exchange are the hallmarks of the modern global economy. This reminds us that we alone do not determine our destiny, no matter how powerful we are. The United States has never flourished in isolation and cannot possibly today, precisely because of the global nature of science, technology, and other knowledge-intensive activities that drive economic activity. Our Founding Fathers knew this when they reached out to Europe—even a former foe in Great Britain—to exchange ideas as well as products soon after the new nation was formed.

Why has America been such a magnet for scientists? I am a scientist–immigrant to this country, but when I left Australia in the late 1970s, I frankly did not think of myself this way. Immigrants, I then imagined, were people who were fleeing some adversity and seeking a better life in America. But in my case—as in so many thousands of others—it was not a matter of running away from something such as economic, political, or geographic adversity. Far from it, because Australia is a great country in which to live and work. Rather, it was a matter of what I was running toward. I came to the United States because of science: The opportunity in the United States to do science of the type I was interested in—basic research at the very highest possible level—and to be among others with similar interests, ambitions, and capabilities were the attractions.

Most of the scientist–immigrants to this country have come for these same reasons—because of the open U.S. culture that rewards excellence and does not impose restrictions based on prior connections, social status, ethnicity, or national origin. When I arrived here, the scientific staff of the Cold Spring Harbor Laboratory of 1979 numbered little more than 100, and those not born here were more likely to have come from Great Britain than any other nation. Today's staff of about 600 scientists is far more diverse, reflecting infusions of talent from all over the world, including from the nations of Asia. This is a natural progression as countries strive to catch up to the United States to propel their economies, feed their peoples, and enhance the lives of their citizens. Although the United States should appropriately be cognizant of threats of scientific and economic espionage, scientific cooperation will more likely generate trust and understanding. Moreover, we have to be careful that the current anti-immigrant rhetoric does not create a climate that causes people of talent to think twice about coming to America.

Globalization of Science

About a dozen years ago, CSHL started a science conference program in Suzhou, China, that parallels the successful meetings program that has existed on our Cold Spring Harbor campus since 1933. These international meetings have been very successful, attracting scientists from many

countries to exchange ideas in the biological, agricultural, and medical sciences. The focus is mainly on basic science, but issues such as genetically modified food have been discussed at an international level in Suzhou. In 2015, we started a DNA Learning Center for laboratory-based science education for middle and high schools in Suzhou, as part of the international reach of DNA Learning Centers that started 30 years ago in Cold Spring Harbor. Indeed, DNA Learning Centers are now in many countries and in many states of the United States. About four years ago, we began assisting the teaching of advanced science at a school in Beijing, similar to the programs we have at many high schools on Long Island and in New York City. Part of the purpose of these combined efforts is to expose students, teachers, and scientists in China and Asia in general to the culture of the scientific enterprise that has been so successful in the United States.

This past year I spent several weeks in China, and this and previous trips have allowed me to reflect on its progress and think about the future of both of our nations. As I stressed in several of my talks and discussions in Beijing, Guangzhou, and Suzhou, science is a global enterprise. This means that progress made anywhere in the world will soon spread far and wide. In most aspects of science, there are no borders. What someone learns about human disease or crop yield in China or in the United States will benefit people the world over. There are of course important exceptions, such as science and technology that directly impacts national security. But in most cases, scientific results are published and made available to everybody rather than kept under wraps. This is a mission that Cold Spring Harbor Laboratory has pursued with remarkable success for more than a century, most recently in its innovation of bioRxiv, which has become the preprint server of choice, globally, for researchers in the life sciences.

China has a population four times ours and one-fifth of the world's, and it stands to reason that a commensurate portion of the world's brightest minds are Chinese. Despite the fact, however, that China is annually graduating more than twice the number of new bachelor's degree holders than either the European Union or the United States, large numbers of Chinese students—many of the best—still want to come to the United States to pursue higher education. Even as Chinese R&D spending approaches that of the United States in absolute terms—$408 billion versus $495 billion in 2015—America spends five times more on a per-person basis—$1532 versus $292. Opportunities for young scientists in America remain unparalleled. Many Chinese scientists who come choose to stay here because they are recruited by major U.S. research universities. Increasingly, however, others are returning to continue their career in China. If China is to fully participate in the worldwide open exchange of scientific research (both basic and applied), full and open protection of intellectual property should become its national policy. Only then can China begin to fulfill its mission to integrate into the international scientific enterprise and global economy.

Think Boldly

Many institutions have tried to figure out why some are more productive than others. To me, it comes down to a small number of factors, beginning with the fostering of a youth culture. We have been able to show that by committing significant resources to the most promising young people—of whatever nationality—we continually reap tremendous benefits in the form of both creativity and productivity. Other factors include the freedom of faculty to pursue research in any area; stressing basic science as the driver for innovation; creating an environment in which collaboration and interactions are encouraged; and providing a physical environment that supports researchers and their families (good housing, local schools, and childcare are all essential).

Within the institution, flat administrative organization is also an element of success—having an administration that serves the science, rather than the other way around. It is important, too, to provide a full range of shared scientific resources, so that even the newest faculty member, or

the least experienced, has ready access to the most sophisticated (and typically very costly) technology and equipment. Finally, consistency of financial support is vital. A program of original, high-risk, high-reward research can take many years to play out. Those who undertake these risks in the formative years of their career should be comfortable in the knowledge that the institution has their back.

As globalization, ease of international transport, the digital revolution, and rising economic activity provide opportunities for other nations to challenge established leaders in science and technology, the United States must continually reexamine its own role in the world. As we do, I urge above all that we continue to think boldly. It would be a grave error to close ourselves off from the rest of the world or to cease cooperating with rising competitors. The very basis of our strength, as I have argued, is our culture of openness and inclusiveness, and that culture that defines American enterprise, including science, is very difficult to reproduce elsewhere. We should continue to welcome the best minds in the world to work and study here—for our benefit and for the world's.

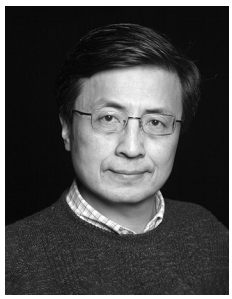
Bruce Stillman, Ph.D., F.R.S.
President and Chief Executive Officer

Highlights of the Year

Research

More than 600 scientists working in Cold Spring Harbor Laboratory's 50-plus laboratory groups contributed to research that in 2017 was published in the world's major research journals. Their efforts reflect the full spectrum of CSHL's scientific activity in Cancer, Neuroscience, Plant Biology, Quantitative Biology, and Genomics. It is impossible in a small space to adequately represent the scope of this work. The following is a sampling of important findings.

Neuronal Types Are Defined by Genes that Shape Their Communication Patterns



J. Huang

Scientists led by Josh Huang have published an important discovery about the molecular-genetic basis of neuronal cell types. Neurons are the basic building blocks that wire up brain circuits supporting mental activities and behavior. The study, which involves sophisticated computational analysis of the messages transcribed from genes that are active in a neuron, points to patterns of cell-to-cell communication as the core feature that makes possible rigorous distinctions among neuron types across the mouse brain.

The team likens their discovery to the way communication styles and patterns enable people to learn important, definitive things about others; to a significant degree, we are defined by the circle of people with whom we communicate. Using six genetically identifiable types of cortical inhibitory neurons that all release the neurotransmitter GABA, the team sought to discover factors capable of distinguishing the core molecular features of the neurons. Using a high-resolution RNA-sequencing method to identify genes that are expressed in individual neurons, coupled with computational analysis tools developed by Jesse Gillis' group, the team searched for families of genes whose activity exhibited characteristic patterns in each neuron type.

Out of more than 600 gene families, the team found about 40 families whose activity patterns could be used to distinguish the six groups of cells. Surprisingly, Huang says, these cell-defining features fell into just six functional categories of gene families, all of which are vital for cell-to-cell communications. The findings should help scientists sort out the bewildering array of neurons that are intertwined in the brain and come up with a code for how they connect and interact.

Dopamine Neurons Factor Ambiguity into Predictions We Depend on

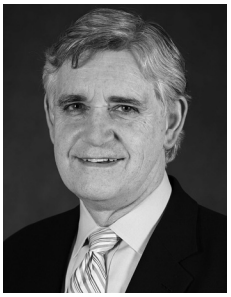
A. Kepecs

Our evolutionary success depends upon our ability to learn and adapt to new conditions, especially when they are changing rapidly. But to make winning decisions, we cannot rely on hardwired instructions. Our success depends on our ability to learn from both successes and failures. Adam Kepecs and his team have been studying how the brain learns, and in recent research, learning how the brain operates in situations where the incoming information is ambiguous.

How does the brain make decisions when the inputs are uncertain? Dopamine-releasing neurons are involved in producing critical teaching signals for the brain, somehow weighing ambiguity—perhaps by reviewing how successfully past experiences guided a new decision. Kepecs' team concludes that dopamine neurons compare predicted outcomes to actual outcomes and send the discrepancy between these as an "error feedback" to many other parts of the brain. This kind of reinforcement learning has been incorporated into many types of artificial intelligence.

The team's research shows that this process in neurons is more complex than previously thought. Their revised model, based on mathematical insight, generates an estimate of the probability that a given choice is correct. Kepecs calls this a measure of the degree of confidence about the decision—in essence, a prediction of accuracy.

New Images of ORC Complex Help Solve Three Biological Mysteries



B. Stillman

L. Joshua-Tor

A ring-shaped protein complex called ORC, or origin recognition complex, performs the first step in the precisely choreographed genome-replication dance that ensures that before cells divide, the genome is duplicated by DNA replication once and only once. Discovered by Bruce Stillman and colleagues in 1991, ORC continues to fascinate scientists, partly because of its multiple functions in the cells and partly because of the difficulty in obtaining images of it at atomic resolution.

Using X-ray crystallography and cryo-EM (electron microscopy), a team led by Leemor Joshua-Tor, in collaboration with Stillman, this year obtained images of human ORC in its active mode at unprecedented resolution. ORC complexes self-assemble in the cell nucleus and bind at specific spots called start sites or origins along the DNA double helix. In human cells, ORC assembles at thousands of origin sites across the entire genome to form an initial configuration called the pre-replication complex, or pre-RC. Each complex requires fuel, which is supplied by adenosine triphosphate (ATP).

In ORC's active phase, the researchers showed that a subassembly containing five ORC subunits engages multiple ATP molecules and forms a partial ring-shaped complex. ATP is also used to recruit another protein component called CDC6, transforming the open ring into a closed ring. By this time, the multipart assembly has engaged and bound to the DNA double helix, which passes through the center of the ring like a bolt through the center of a nut.

The new images help resolve three outstanding mysteries: how DNA binds with ORC, how the ATP fuel is used, and how mutations in ORC complex proteins give rise to a human disorder called Meier–Gorlin syndrome.

Using CRISPR Scissors to Vary Traits in Tomato

Z. Lippman

Zach Lippman and colleagues have used CRISPR-Cas9 technology to rapidly generate variants of the tomato plant that display a broad continuum of three separate, agriculturally important traits: fruit size, branching architecture, and overall plant shape. All are major components in determining how much a plant will yield. The method is designed to work in all food, feed, and fuel crops, including staples like rice, maize, sorghum, and wheat.

Using CRISPR to create different sets of mutations in a gene promoter called SICLV3 (and in several other promoters), the team was able to introduce a wide range of variations in the number of floral organs and locules (gelatinous seed compartments) in tomato plants. The effect is analogous to turning a dimmer switch to vary light levels over a continuous range. In this instance, as the gene's activity declines, the number of flower petals increases, as does the number of seed compartments in the resulting fruit—and, hence, fruit size increases.

All of these effects can be traced to changes in stem cell number in the plant's stem cell reservoir, called the meristem. Traditional genetic breeding involves great time and effort to adapt beneficial variants of relevant genes to the best varieties, which must continuously be improved every year.

The new approach bypasses this constraint by directly generating, and selecting for, the most desirable variants controlling gene activity in the context of other natural mutations that benefit the quality and quantity of fruit in plants.

A Defender of the Genome When It Is Naked

Our genomes are minefields, studded with potentially damaging DNA sequences over which hundreds of thousands of sentries stand guard. Called epigenetic marks, these sentries attach to the double helix at such spots and prevent the underlying DNA sequences from springing into

destructive action. About half the human genome is composed of these damaging sequences. They are where ancient viruses and parasitic elements called transposons and retrotransposons have incorporated themselves over the long course of evolution. These genetic elements need to be kept silent to maintain integrity of the genome.

R. Martienssen

It's astonishing, then, to consider that during two of the most crucial processes in the life cycle, the sentries are removed, leaving the genome naked. The defenders are quickly restored, but only after an interval in which the epigenetic slate is wiped clean. The short period when the genome is naked of the epigenetic marks allows the new embryo to reset and gain its own character, different from the marks that were inherited on the chromosomes from the parents. A team led by Rob Martienssen has now uncovered the existence of what might be considered emergency replacements for the sentries—troops pressed into service across the genome only during these curiously undefended moments.

These defenders are protecting the genome in mammalian embryos, at the very early stages of development before the embryos are implanted in the wall of the maternal uterus. The newly identified defenders are RNA fragments 18 and 22 nucleotides in length. These fragments are perfect complements of sequences in retrotransposons that must be engaged in order for the genomic parasites to be activated.

Martienssen's team thinks the cell is deliberately chopping up full-length transfer RNAs (tRNAs) into smaller fragments precisely because both tRNAs and the fragments cut from them recognize a binding site on retrotransposons that is essential for them to become active. This means the small, tRNA-derived fragments would be able to occupy that site and inhibit retrotransposon replication and mobility, thereby protecting the genome while it is being reset with the new epigenetic marks. This could be one way the genomes of mammals have tolerated vast numbers of transposons and other parasitic elements.

The First Cell-Type Census of Mouse Brains

P. Osten

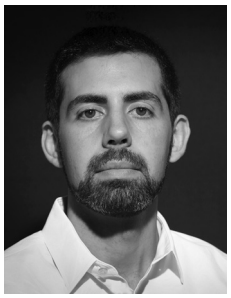
A team led by Pavel Osten has mobilized advanced imaging and computational methods to comprehensively map, or "count," the total populations of specific types of cells throughout the mouse brain. Their "qBrain" (quantitative brain) demonstration revealed that contrary to expectations, the numbers and ratios of three major inhibitory cell types vary in a stereotypical way across different parts of the mouse cortex.

This implies that different cortical areas—for instance, those involving cognition versus those involving perception of sensory stimuli—have evolved to tailor their local circuits to specific brain functions. It is also surprising that although male and female brains did not differ in cell counts in cortical regions, the study identified 11 subcortical areas with sex-specific differences. Strikingly, despite the overall tendency for male brains to be larger, 10 of these regions had more modulatory neurons in females than in males. This shows that, in most respects, there are more cells that modulate signals and exert temporal control in areas regulating reproductive, social, and parenting behaviors in females than in males.

qBrain is built on an automated technology platform that will be used to perform similar analyses of other mammalian brains, from prairie voles to marmoset monkeys and humans. This will enable unprecedented cross-species comparisons.

A More Complex Relation between Chromosome Imbalance and Cancer

More than a century ago, a German-born scientist, Theodor Boveri, reasoned that having the wrong number of chromosomes could cause cells to grow uncontrollably and become the seeds of cancerous tumors. We now know that 90% of solid tumors and 75% of blood cancers have aneuploidy, or abnormal chromosome numbers. But new research from Jason Sheltzer's group suggests that the relationship between aneuploidy and cancer is more complex than previously believed.



J. Sheltzer

They discovered that cells with a single extra chromosome that had been primed to become cancerous actually grew more slowly and formed smaller tumors than similarly primed cells with normal chromosome counts. After a period of weeks, however, the slower growers experienced explosions of growth and displayed pronounced genetic instability.

Sheltzer suspects they rapidly evolved new genomic mutations that enhanced their ability to survive with an extra chromosome. Such rapid evolution in premalignant cells may account not only for their cancerous transformation, but it may also help explain characteristics seen in metastatic cancer cells such as gaining the ability to move to new locations in the body and to resist toxic chemotherapy.

A Missing Protein that Can Drive Prostate Cancer Progression

L. Trotman

A protein called PTEN is one of the body's tumor suppressors. Mutations in the gene encoding it are commonly found in many different types of cancer. Yet some cancer patients show low levels of the PTEN protein even though their *PTEN* genes are normal. Lloyd Trotman and colleagues have discovered that this may be due to defects in a protein called Importin-11, which transports PTEN into the cell nucleus, sheltering PTEN from proteins in the cytoplasm that would otherwise target it for degradation.

Specifically, they demonstrated that loss of Importin-11 may destabilize PTEN, leading to the development of lung, prostate, and other cancers. Mutations in the gene encoding Importin-11 have been identified in human cancers, and Trotman and colleagues found that tumors from lung cancer patients lacking Importin-11 tended to show low PTEN levels as well. The researchers estimate that loss of Importin-11 may account for the loss of PTEN in approximately one-third of lung cancer patients lacking this key anticancer protein.

In prostate cancer, loss of Importin-11 predicted disease relapse and metastasis in patients who had their prostate removed. Trotman's results suggest that Importin-11 is the "Achilles' heel" of the protein-based machinery called the ubiquitination system, which maintains the correct levels of PTEN inside cells.

Fibroblast Varieties May Help Explain Why Pancreatic Cancer Is So Hard to Treat

D. Tuveson

Why are pancreatic tumors so resistant to treatment? One reason is that the wound-like tissue surrounding tumors, called stroma, is much more dense than stromal tissue surrounding other, more treatable tumor types. Stromal tissue is also believed to contain factors that aid tumor survival and growth. In pancreatic cancer, its density is thought to be a factor in preventing cancer-killing drugs from reaching the tumor.

David Tuveson's team used pancreatic organoid technology to learn something important about the problem of stroma in pancreas cancer. For the first time, their pancreatic organoids were "co-cultured" with one component of the stroma in which human tumors grow. The result was a more realistic rendering of what happens in patients.

The additional factor was CAFs (cancer-associated fibroblasts), which act like factories in the tumor, producing connective tissue. The team discovered that there are at least two varieties of CAFs in pancreatic cancer. Each seems to be involved in different ways. One fibroblast subtype produced a protein called αSMA; it contributes to the formation of dense stroma. The other, which secretes immune factor IL-6, has been separately linked to cancer cell proliferation.

Discovery of the heterogeneity of the fibroblast population in pancreas cancer opens up the possibility of selectively targeting these populations to make treatments more effective.

An Epigenetic Explanation of Metastasis

C. Vakoc

Although cancer is well understood to have genetic causes, the factors responsible for its ability to spread have eluded scientists. Results obtained by a team led by Chris Vakoc, which included researchers in David Tuveson's lab, make a case for metastasis to have epigenetic roots—that is, critical changes in the chemical modifications on the chromosomes that do not change the underlying DNA sequence. The team used organoid technology to compare normal pancreatic ductal cells in mice with cells from the same animals in three distinct stages: premalignant, malignant, and metastatic.

Compared with cells in noncancerous and tumor-derived organoids, those in metastatic ones displayed an extraordinary number of alterations in regions called enhancers. These genome elements are present in all cells and are used by cells to turn on genes. "We show that to metastasize, the cell has to change, in effect, its whole telecommunications network—its enhancers are being reprogrammed," Vakoc says.

The team tracked down the molecular agent whose increased activity causes the reprogramming. Called FOXA1, it is typically active early in a cell's life, but dormant later on. In cancer, FOXA1 activity enables a cell to return to a developmentally primitive state—one in which the pancreas is being formed, and cells are multiplying and moving around, assuming the positions that characterize the maturing organ.

The new evidence of cancer cells acquiring the ability to spread by "remembering" a developmental program dormant since their earliest days "means that every cell, in a sense, is like a loaded gun," Vakoc says.

Improved Genome Reveals Maize's Remarkable Adaptive Potential

D. Ware

A new, much more detailed reference genome for maize, or corn, was published this year by a team led by Doreen Ware and her colleagues at CSHL and around the globe. In its accounting of the sequence of nucleotides in the plant's 10 chromosomes, the new version, which was obtained for a small fraction of the cost of the first reference genome for the plant, published in 2009, helps us understand as never before why maize is the most productive and widely grown crop in the world.

Among many other things, the new sequence reveals that maize individuals are much less alike genomically than people are. This reflects its remarkable flexibility. This flexibility not only helps explain why maize has been so successful since its adaptation by agriculturalists thousands of years ago, but also bodes well for its ability to grow in new places as the earth's climate changes and for increasing the plant's productivity and environmental sustainability in the United States and abroad.

The new research demonstrates that in trying to determine what possibilities are available to a plant when adapting to new or changing conditions, it is just as much the context in which genes are activated—or silenced—as the identity of the genes themselves that determines what the total set of genes enables a plant to do.

Public and Private Support

The year 2017 vividly demonstrated how CSHL is changing the landscape of biology. New York State Governor Andrew Cuomo and other state and local officials joined CSHL trustees and leadership to break ground on the renovation of the Demerec Laboratory, the focal point for a new CSHL research initiative to establish linkages between the development of cancer and nutrition, obesity, and metabolism. The $75 million effort is possible thanks to funds raised by the 125th Anniversary Capital Campaign and a grant from New York State.

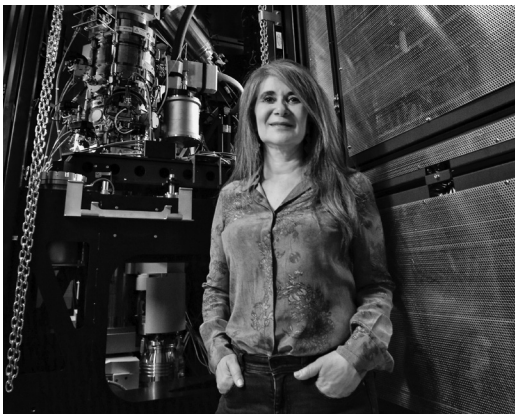
Breaking ground on the renovation of the Demerec laboratory.
Bruce Stillman is *third from left* and Governor Andrew Cuomo is in the *center*

CSHL's historic commitment to widespread and effective dissemination of scientific knowledge was enhanced by the growing momentum of bioRxiv, a bold Laboratory initiative creating the first preprint server for the life sciences.

In its fourth year of operation, bioRxiv is transforming how life scientists communicate. Preprints are complete but unpublished manuscripts of research papers, distributed online by their authors for discussion by the scientific community before consideration by peer-reviewed journals. Submissions have soared, more than doubling in 2017 to a total of some 17,000 manuscripts from researchers in more than 100 countries.

In recognition of bioRxiv's role in accelerating science, generous support was received in May from the Chan Zuckerberg Initiative, which will make possible continued expansion and a variety of technology upgrades and innovations.

With support from Mercer Family Foundation, the Laboratory now has a state-of-the-art Cryo-EM Facility designed for optimal operation of an FEI/Thermo Fisher Titan Krios G3. Led by Professors Leemor Joshua-Tor and Hiro Furukawa, and managed by Dennis Thomas, this facility is for biologists who seek to define the detailed structure of molecules and allows researchers to obtain near-atomic-level 3D images. The CSHL Shared Resource for Animal Imaging &



Leemor Joshua-Tor with the Thermo Fisher Titan Krios G3

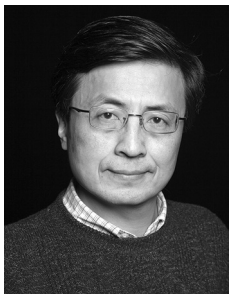
H. Furukawa

Tissue Imaging, which provides access to the most advanced noninvasive imaging modalities, tissue imaging, and pathology, was opened in 2017. The new facility is critical to the work of CSHL's National Cancer Institute–designated Cancer Center in testing potential cancer therapeutics and was funded by David H. Koch, in addition to an anonymous donor, and funds from New York State. Researchers can now visualize a broad range of tumor-associated parameters without invasive procedures, obtain high-quality tissue sections, and access extensive pathology services.

With $50 million over 5 years committed in grant funding from the National Institute of Mental Health (NIMH), CSHL is proud to be part of the BRAIN Initiative Cell Census Network (BICCN). Under this initiative, led by Professor Josh Huang, CSHL this year established a Center and a "Collaboratory" for the Mouse Brain Cell Atlas. Many of CSHL's neuroscience faculty have secured BRAIN Initiative grants, including Jesse Gillis, Pavel Osten, Partha Mitra, and Tony Zador. Additional Brain Initiative funding was secured to support Pavel Osten, Tony Zador, Adam Kepecs, and Bo Li.

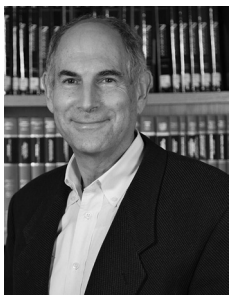
The Wellcome Trust and the Simons Foundation are supporting the International Brain Lab, an initiative including CSHL investigators that seeks to develop theories of how the brain works by focusing on a single behavior shared by all animals: foraging. Among the executive leaders of the global program are Associate Professor Anne Churchland and Professor Tony Zador.

CSHL's Annual Fund raised a record $7 million in unrestricted funds through many successful events, including the Double Helix Medals Dinner honoring Tom Brokaw and Helen and Charles Dolan; the Women's Partnership for Science, celebrating CSHL Association honorary director Freddie Staller; and the Golf Tournament, applauding Mark Hamer.



J. Huang

J. Berger-Sweeney



S. Weisbrod

T. Brokaw

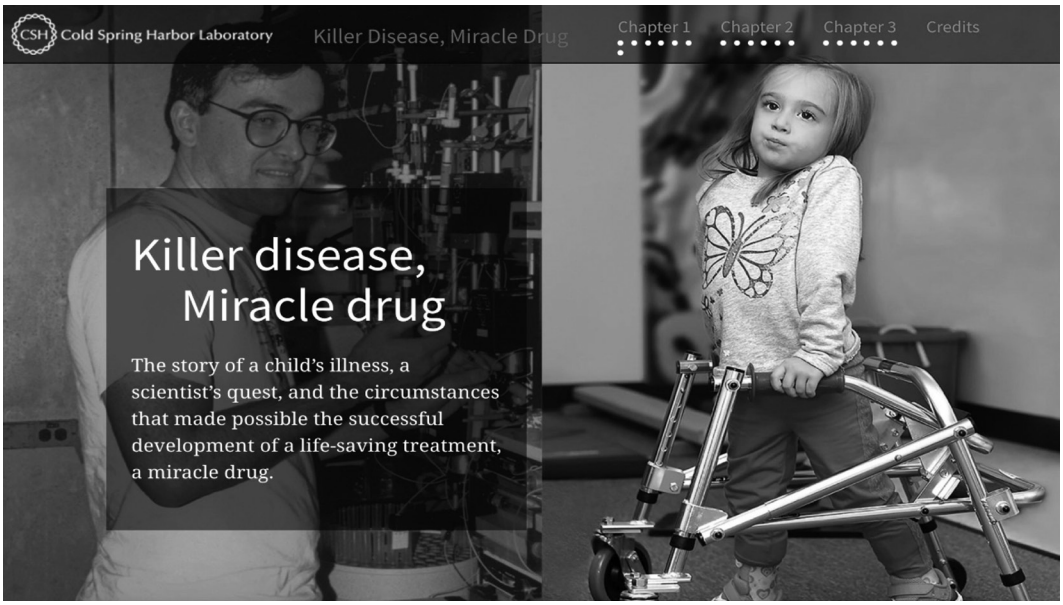
Board of Trustees

The Board of Trustees welcomed two new members: Joanne Berger-Sweeney, Ph.D., President of Trinity College, and Stuart Weisbrod, Ph.D., Chief Investment Officer of Iguana Healthcare Partners, a healthcare investment fund focusing on public equities.

Berger-Sweeney became President of Trinity College in 2014 after serving as Tufts University Dean of the School of Arts and Sciences. Prior to this, she was an associate dean and faculty member at Wellesley College. Berger-Sweeney received an undergraduate degree in psychobiology from Wellesley College and an M.P.H. in environmental health sciences from the University of California, Berkeley. She has a Ph.D. in neurotoxicology from the Johns Hopkins School of Public Health, where she did the proof-of-concept work on Razadyne, the second-most-used Alzheimer's drug in the world.

Helen and Charles Dolan

L. Joshua-Tor



Weisbrod first came to CSHL as a postdoctoral fellow after receiving a Ph.D. in biochemistry from Princeton University. He then pursued an MBA at Columbia University and began a biotech investment career capped by the founding of Merlin Biomed and later Iguana Healthcare Partners.

Research Faculty Awards

Professor and Howard Hughes Medical Institute Investigator Leemor Joshua-Tor was elected to the National Academy of Sciences as one of 84 new members and 21 foreign associates recognized for "their distinguished and continuing achievements in original research." She also joined the ranks of the American Academy of Arts and Sciences.

M. Egeblad

Joshua-Tor's laboratory studies the molecular basis of nucleic acid regulatory processes—RNA interference (RNAi), and DNA replication in particular. They use the tools of structural biology, biochemistry, and biophysics to study proteins and protein complexes associated with these processes to elucidate how they work.

The first ever drug to treat a lethal childhood disease called spinal muscular atrophy, Spinraza®, won the 2017 Galien Prize for Biotechnology Breakthrough of the Year. Marketed by Biogen, the drug was developed by Ionis Pharmaceuticals Inc. with technology licensed from and in collaboration with Professor Adrian Krainer's laboratory. Adrian was also named Inventor of the Year by the New York Intellectual Property Law Association.

Associate Professor Mikala Egeblad won the Pershing Square Sohn Cancer Prize, which will support her work to understand the relationship between chronic inflammation and the metastatic recurrence of breast cancer. Her team will explore the role of neutrophils, a specific type of immune cell involved in the awakening of dormant cancer cells. Neutrophils can form neutrophil extracellular traps (NETs) as part of the body's reaction to inflammation. By targeting NETs, Egeblad and her team hope to be able to prevent cancer recurrence in certain cases.

D. Spector

David L. Spector, Ph.D., the Laboratory's Director of Research, was named a fellow of the American Society for Cell Biology (ASCB). A member of the CSHL faculty since 1985, Spector is a pioneer in advancing our understanding of the inner workings of the cell nucleus, researching the organization and regulation of gene expression in living cells. He is the first CSHL faculty member to receive this honor.

In recognition of her efforts to promote and mentor women in neuroscience, Associate Professor Anne Churchland was honored with the Louise Hanson Marshall Special Recognition Award at the Society for Neuroscience's 2017 Annual Meeting in November. Each year, the award goes to "an individual who has significantly promoted the professional development of women in neuroscience through teaching, organizational leadership, public advocacy, or other efforts that are not necessarily research-related."

A. Churchland

Churchland has garnered attention from the neuroscience community and beyond for her pioneering website, Anne's List, aimed at alleviating bias against women in science—specifically, as reflected in the paucity of invited female speakers at scientific conferences. Anne's List gathers female scientists' names, research topics, and seniority levels in order to help conference organizers easily identify women whom they might invite to speak on a particular subject. Churchland is also a faculty advisor for the group Women in Science & Engineering (WiSE) at CSHL. Her work on Anne's List helped inspire CSHL WiSE to team up with CSHL's Meetings & Courses division to create a larger Women in Biology Speakers List.

T. Engel

Promotions and New Hires

The Laboratory welcomed Assistant Professor Tatiana Engel to CSHL's Swartz Center for Computational Neuroscience, where she is focused on the dynamics of neural circuits.

Monn Monn Myat, Ph.D., is the new Associate Dean of the Watson School of Biological Sciences.

M.M. Myat

CSHL-Northwell Affiliation 2016–2017 Progress

During the second year of the strategic partnership between CSHL and Northwell Health System, we established an infrastructure that allows clinicians and basic scientists to easily share patient samples. Through these efforts, CSHL received tissue samples from 140 Northwell Health patients. These samples bring clinical data into the laboratory, and they are being used to identify genes that drive cancer and to test potential new therapies. The affiliation is also expanding translational research to make a direct impact on patient care. This means bringing discoveries out of the laboratory and into the clinic.

We want to ensure that the next generation of doctors has the tools and knowledge they need to bridge the gap between basic science and patient care. This includes expanding upon and developing new education initiatives that promise to transform the way young doctors are trained in clinical research. Medical students from the Donald and Barbara Zucker School of Medicine at Hofstra/Northwell continue to participate in research during their summers, and next year we hope to bring on a Translational Research Fellow through the National Resident Matching Program.

The original 2015 agreement was amended to include a small component for neuroscience. Now, the affiliation will fund high-impact projects to increase translational research in both neuroscience and cancer. Over the past year, the Joint Steering Committee has funded eight new collaborative research proposals for a total of 28 ongoing projects.

The affiliation has sponsored international cancer conferences and meetings held at CSHL, including Automatic Pathology and Making Oxidative Chemotherapy Less Toxic. CSHL was awarded a subcontract by Leidos Biomedical Research to lead a Cancer Model Development Center for multiple cancers. Drs. David Tuveson and David Spector will lead the international effort with Dr. Hans Clevers of the Hubrecht Institute, Drs. Aldo Scarpa and Vincenzo Corbo of the ARC-Net Centre for Applied Research on Cancer at the University of Verona, Italy, and Dr. James

R. Barakat

Crawford of Northwell Health and Dr. Peter Gregersen of Northwell Health's Feinstein Institute for Medical Research.

Northwell Health has continued efforts to build out a Phase 1 Experimental Therapeutics Unit to be led by Dr. Robert Maki, who holds a joint appointment at CSHL and Northwell.

The affiliation announced the hiring of Dr. Richard Barakat, Physician-in-Chief and Director of the Northwell Health Cancer Institute, Senior Vice President of the Cancer Service Line and Professor of Obstetrics and Gynecology at the Zucker School of Medicine at Hofstra/Northwell.

Business Development & Technology Transfer

Engaging industry and investor partners is a means to bring CSHL's innovation to the world while providing the business with resources, capital, and philanthropic funding, as well as a competitive advantage in the commercialization of an innovation if it is intellectual property. We look to these industry partners not only for investments, but also for consultative business development guidance in an ever-more-complicated business climate. A goal is to find excellent sponsored research partners, and the number and value of these engagements has increased significantly in the past 3 years. We have moved from sourcing, negotiating, and signing one or two arrangements a year to completing eight to 10 per year, involving a variety of industry partners and CSHL labs.

This year we evolved our approach to engaging industry and investors. In cooperation with the Development Department and the Meetings & Courses Program, we have introduced potential industry partners and investors to the Banbury Center, Meetings & Courses Program sponsorship opportunities, donor engagement, and postdoc and graduate student career discussions, as well as research collaborations.

Dr. R.K. Narayan, Ph.D., joined the team as Director, Technology Transfer, with an initial emphasis on operations and compliance matters, and as a lead for technology transfer matters in neuroscience. We also added to the Executives in Residence and Entrepreneurs cadre of business executives, who advise faculty on commercial aspects of their work and serve on the Boards of some of CSHL's spin-out companies: Kate Delgado, CSHL Entrepreneur in Residence, and Peter Young, CSHL Executive in Residence.

In 2017, while focusing on the management of existing agreements and assets, compliance and risk mitigation, we concluded the following:

License/option agreements:	3
Sponsored research agreements:	9
Material transfer agreements:	142
New patent filings:	12

Education Programs

Meetings & Courses Program

CSHL Meetings this year attracted 7,300 participants to the main campus. The 82nd Cold Spring Harbor Symposium, Chromosome Segregation & Structure, addressed the enormous progress in this field, attracting almost 300 participants, including many of the world's leading chromosome biologists. The year saw the continuation of many successful annual and biennial meetings, as well as the introduction of several new meetings. The meetings program is supported by grants from the National Institutes of Health, the National Science Foundation, and the newly invigorated Corporate Sponsor Program.

The Cold Spring Harbor Asia (CSHA) conference program held 17 scientific conferences in Suzhou, China, attracting more than 3,250 scientists. CSHA's scientific program, which includes symposia and meetings and occasional Banbury-style discussion meetings, is designed for scientists from the Asia/Pacific region, who make up more than 80% of attendance. The program is supported with a major sustaining grant from the Suzhou Industrial Park, where the program is headquartered.

The Courses program has played an important role in the history of molecular biology and life science, propagating important new techniques, methods, and ideas among scientists at all career stages. Covering a diverse array of topics in molecular biology, neurobiology, structural studies, and bioinformatics, 750 instructors, lecturers, and assistants have come to teach at CSHL from universities, medical schools, research institutes, and companies around the world. In 2017, 600 trainees—advanced graduate students, postdocs, and faculty—attended courses lasting from 1 to 3 weeks.

The Courses program relies on grants and foundation support, including major support from the Helmsley Charitable Trust, Howard Hughes Medical Institute, National Institutes of Health, and National Science Foundation. The Helmsley Interdisciplinary Fellowship Fund provided major funding to 125 scientists to participate in CSHL courses outside their primary disciplines. The Courses also benefit from the loan of equipment, reagents, and technical support from many companies, whose support is indispensable to ensure that the program remains cutting-edge.

Watson School of Biological Sciences

In 2017, the Watson School welcomed its 19th incoming class and graduated its 14th. The achievements of the graduate program continued to grow. The quality of scientific publications produced by the School's students remained highly impressive. Watson School students continued to graduate considerably faster than students in comparable Ph.D.-granting institutions and demonstrated an ability to secure excellent jobs.

Twenty-six graduates have now secured tenure-track faculty positions and are receiving federal grants and publishing papers as independent researchers. Eight of them have already been promoted to associate professor, and Zachary Lippman became the first to be promoted to full professor, here at CSHL. The School's graduates have also moved into influential positions in administration, publishing, consulting, and industry.

At the 2017 graduation ceremony, eight WSBS students were awarded Ph.D. degrees, bringing the total since the School's inception to 98.

2009 Nobel laureate Carol Greider received an honorary Doctor of Science degree at WSBS graduation. She earned her Ph.D. at the University of California, Berkeley, where as a graduate student in 1984, working with Dr. Elizabeth Blackburn, she discovered telomerase, an enzyme that maintains telomeres—the "caps" at the end of chromosomes. In 1988 Dr. Greider came to Cold Spring Harbor Laboratory, where, as the Lab's second CSHL Fellow, she cloned and characterized the RNA component of telomerase. In 1990 she was appointed an Assistant Investigator and in 1994 an Investigator. In 1997 Dr. Greider moved her laboratory to the Johns Hopkins University School of Medicine. Today she is Daniel Nathans Professor and Director of Molecular Biology and Genetics at the university.

During the year, scientific papers published by students of the School appeared in major journals, bringing the cumulative total to more than 365. Current and former students won prestigious and highly competitive scholarships and fellowships, as in past years. In August, the WSBS welcomed eight new students. Members of the Class of 2017 were selected from more than 200 applicants. Other new graduate students entered as visitors from other institutions, including 10 from Stony Brook University.

2017 WSBS graduation

From June through August, 20 undergraduates from around the United States, as well as China, Greece, and Pakistan, had the remarkable opportunity to perform advanced research in the laboratory of a CSHL faculty member. This immersive experience brought intellectual as well as social rewards for the participants, as in past years. The URP (Undergraduate Research Program), along with the equally innovative Partners for the Future program, which brings gifted local high school students to CSHL labs for hands-on research experience, are run and managed by the Watson School.

Banbury Center

The year 2017 was one of transition for the Banbury Center, the Laboratory's science policy think tank. Dr. Jan Witkowski, the founding director, handed the baton to Dr. Rebecca Leshan after 30 remarkable years of cultivating critical scientific discourse at the Center. In his speech, "How Scientists Work," at Banbury's 1977 dedication ceremony, Francis Crick pointed to small meetings as the best way for scientists to share and inspire new ideas and strategies. Activities at Banbury continue to be guided by this concept; its mission is to further scientific knowledge and the well-being of society.

Banbury meetings in 2017 spanned discovery and translational science, public health, policy, education, and innovation, reflecting the ever-growing need for multisector and multidisciplinary engagement at small meetings across a broad range of issues. A total of 536 individuals took part in these Banbury meetings, with 72% marking their first occasion. Participants were drawn from 28 countries on six continents.

In 2017 Banbury continued to attract financial support from across sectors, with more than half drawn from not-for-profit organizations. Five of 2017's meetings built on a strong history in neuroscience and cancer. Public health was the subject of several productive meetings, including Maximizing Impact of New HIV Prevention Technologies in Sub-Saharan Africa and Protective Immunity & Vaccines for Lyme Disease. A report on next-generation Lyme disease diagnostics was published in *Clinical Infectious Disease*, based on a 2016 Lyme disease meeting at Banbury.

NLR17 group meeting at Banbury

DNA Learning Center

The DNA Learning Center continues to spread its hands-on approach to teaching biology and genome science to students across the globe. A licensed DNA Learning Center established in China in 2014 at Beijing No. 166 Schools now extends to include middle school as well as high school students. In 2017, a teacher-training program and a DNA barcoding citizen science project for high school students were also started there.

Barcode Beijing, modeled on student DNA barcoding projects that have succeeded on Long Island and in New York City schools, involves middle/high school students in independent, student-driven research projects that use DNA sequencing to study biodiversity in their own environs. Using a single, standardized chemistry and bioinformatics platform, students explore many aspects of the urban environment—wildlife in homes and parks, species used in commercial products, insect disease vectors, introduced species, and food mislabeling.

Another novel concept that the DNALC has helped to successfully propagate internationally is the Breakthrough Junior Challenge. Begun in 2015, it is a global competition in which precollege students produce short videos explaining an important concept in life sciences, mathematics, or physics. Funded by Mark Zuckerberg and Priscilla Chan and Yuri and Julia Milner, the Junior Challenge is a complement to the prestigious Breakthrough Prize, designed to inspire creative thinking about fundamental concepts in the life sciences, physics, or mathematics. In addition to a $250,000 scholarship, winners receive a DNALC-designed and -equipped $100,000 science lab for their schools.

This year, in fulfillment of an award in the competition won last year by Hillary Diane Andales from the Philippines, the DNALC established a training lab at the Eastern Visayas Campus of the Philippine Science High School and trained local teachers so that all 14 campuses of the Philippine national science high school system can benefit from the new school lab classroom.

In the New York metro area this year, 21,000 students attended labs at Dolan DNA Learning Center, DNALC West, and Harlem DNA Lab. An additional 9,000 students completed labs in school led by DNALC staff, and 1,400 students attended week-long camps. Also this year, 6.4 million visitors accessed DNALC's suite of multimedia resources online, including 4.6 million visits to DNALC websites, nearly one million views of YouTube videos, and more than 815,000 downloads of smartphone/tablet apps, the 3D Brain, Weed to Wonder, and Gene Screen.

Breakthrough lab at the Philippines High School

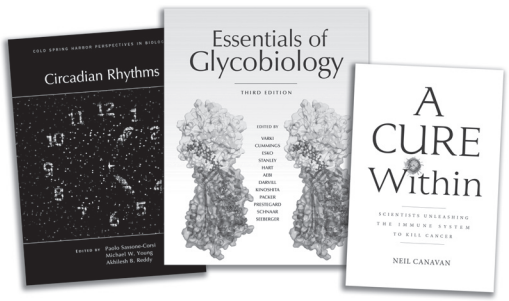
Cold Spring Harbor Laboratory Press

Science is a global enterprise that depends on the timely communication of ideas and results. The Cold Spring Harbor Laboratory Press provides scientists worldwide with authoritative, affordable, and pertinent information to further their research and aid in their career development. The CSHL Press has taken a creative approach to challenges that have attended the rise of digital media.

Its three newest peer-reviewed subscription journals—*Cold Spring Harbor Protocols*, *Cold Spring Harbor Perspectives in Biology*, and *Cold Spring Harbor Perspectives in Medicine*—enjoy considerable success. Each transforms content that in former years might have appeared only in print book form, rendering it in a digital, readily discoverable, and reusable serial form.

The newest journal, *Cold Spring Harbor Molecular Case Studies*, reimagines the traditional case report in medicine to enable open sharing of insights that genomic and molecular analysis bring to the causes and potential treatment of disease.

Genes & Development and *Genome Research* remain at the top of their disciplines among primary research journals. Both titles are in the top 1% of the 8,000 journals ranked in the Science Citation Index.



Some 2017 books

The CSHL Press published 18 new book titles in 2017, adding to its list of more than 200. Highlights including topical volumes on malaria biology (supported by the Flowers Foundation), tissue engineering and regenerative medicine, prion diseases, and the biology of exercise. *A Cure Within: Scientists Unleashing the Immune System to Kill Cancer*, by Neil Canavan, tells the story of immune-oncology pioneers whose work has resulted in two FDA-approved T-cell-based therapies for cancer.

Library & Archives

The 2017 Sydney Brenner Scholar is University of Manchester Professor Matthew Cobb, who is researching the two-decade collaboration between Francis Crick and Sydney Brenner.

Previous Sydney Brenner Scholar Miriam Rich, who is a History of Science graduate student at Harvard University, presented a seminar titled Defects of Development: Embryology & Eugenic

Concepts of Race. This talk was part of the "Conversations on the History of Eugenics," which also included talks and a panel discussion with Library and Archives Executive Director Mila Pollock, historian Elof Carlson, and the Executive Director of the DNA Learning Center, David Micklos.

The CSHL Archives increased this year with several donations. Nobel laureate Carol Greider donated 30 linear feet of laboratory notebooks and laboratory films, which includes the notebooks detailing her Nobel-winning graduate work on telomeres with Elizabeth Blackburn. A donation of an additional 6 linear feet of materials spanning 1950–2008 for the Barbara McClintock Collection came from McClintock's niece, Marjorie Bhavnani. Dr. Winship Herr, former Assistant Director of the Laboratory and Founding Dean of the Watson School of Biological Sciences, donated an additional 6 linear feet of materials to the Winship Herr collection. These personal communications and administrative materials illuminate the effort to gain Ph.D.-granting accreditation from the Board of Regents of the University of the State of New York.

The Archives created an online digital exhibition for the Amar Klar Memorial from archival materials and photographs.

The Library and Archives and Genentech Center for the History of Molecular Biology and Biotechnology, in collaboration with the Laboratory's Meetings & Courses Program, continued its History of Science annual meeting series with 40 Years of mRNA Splicing: From Discovery to Therapy. The meeting, co-organized by Mila Pollock, Phil Sharp, and Joan Steitz, brought together more than 270 of the most important mRNA splicing researchers, including six Nobel laureates. The meeting also brought back our alumni Rich Roberts, Louise Chow, and Richard Gelinas, who told the story of the discovery of mRNA splicing.

Other Library events included a talk on premodern neurosurgery by Dr. Eugene Flamm of Montefiore Medical Center, a celebration of women in STEM fields as explored through books in our Library written by and about women, and a series of seminars, cohosted with the Post Doc Liaison Committee on Perspectives on Science Careers.

40 Years of mRNA Splicing: From Discovery to Therapy meeting

For the second year, the Library offered a multischool journal club, bringing 15 high school students from Long Island to the Laboratory to learn how to search for, identify, examine, and present scientific publications.

Infrastructure

Reconstruction of the Demerec Laboratory began in earnest in 2017. This major project comprised a complete redesign and reconstruction of a ca. 1953 laboratory building. This historic building has been home to some of CSHL's most honored researchers, including Nobelists Barbara McClintock, Alfred Hershey, Rich Roberts, and Carol Greider. New York State contributed $25 million to fund this renovation that involved extensive demolition of the original structure and relocation of many researchers to renovated lab spaces in the Beckman, Jones, and McClintock laboratories. Completion of this project is expected in April 2019.

Modernization of Dolan Hall, a ca. 1991 dormitory, will allow for 60 private rooms with en suite bathrooms. The project was divided into two halves to allow for continued use of the property by Meetings & Courses Program participants while renovations took place. Completion is anticipated in early 2019.

The Olney barn was a ca. 1880 barn originally built alongside the Olney House. Its condition had deteriorated to render the structure unsalvageable. A cosmetically similar replacement structure was built during 2017 to accommodate the need to store and maintain grounds equipment.

The Cryo-EM Facility was constructed this year to house a new, state-of-the-art electron microscope. It is a dedicated isolated facility that has been retrofitted into an existing building, the Beckman Laboratory. This project required complete isolation of the facility from the rest of the building, including isolated ventilation, isolated power, and active electromagnetic shielding.

Community Outreach

CSHL was pleased to work with local area organizations on events that furthered our mutual interests, including the Children's Heart Foundation, Energeia, Friends of TJ, Leukemia and Lymphoma Society, Suffolk County Estate Planning Council, the Lustgarten Foundation, LIA Young Professionals Committee, ALS Association of Greater NY, the Huntington Hospital Board, and the Animal Cancer Foundation.

Cryo-EM facility

Open House

2017 Open House

CSHL welcomed 500+ new friends to campus on June 10 to share the thrill of scientific discovery. Led by the Public Affairs Department, more than 80 CSHL volunteers representing the institution's scientific research and education expertise in genetics and molecular biology helped explain to guests how work at CSHL benefits society.

Guests engaged in hands-on experiments, toured laboratories, and interacted one-on-one with CSHL scientists, learning more about DNA, plant biology, cancer research, neuroscience, and quantitative biology. A special program of half-hour science talks covered topics as diverse as mapping the brain with barcodes, demystifying GMOs, meeting Ötzi the iceman, and how scientists are attacking metastatic breast cancer with new tools found in the genome's "dark matter."

In addition to the campus tours provided at the Open House, the tour guide team of 17 graduate students gave 67 group tours to more than 1300 guests and 16 public tours for more than 250 participants.

First graders from neighboring Cold Spring Harbor School District's Goosehill Primary School and Friends Academy participated in a hands-on science fair consisting of six stations. At each station, the students learned about various scientific principles (from the brain and magnetism to DNA codes and enzymes) through activities and instruction conceived, planned, and led by Watson School graduate students and DNALC teachers. The participants included 120 students accompanied by 10 teachers and more than 100 parents during the 2 days.

Expanding CSHL's capabilities to engage audiences, on-demand and on mobile devices, around the world, the Public Affairs Department launched a technology upgrade to cshl.edu. Together with increased efforts to develop new digital content and leverage social media channels, the Laboratory research and education developments are being communicated as multimedia stories that can be more readily understood by individuals not formally trained in science. The cshl.edu website is the platform for the Lab's external communications programs.

CSHL Public Lectures

March 1: Rob Martienssen, Ph.D., Professor, Cold Spring Harbor Laboratory: *Cocktails & Chromosomes*.

May 24: Adam Kepecs, Ph.D., Professor, Cold Spring Harbor Laboratory: *Cocktails & Chromosomes*.

June 14: Adam Siepel, Ph.D., Professor, Cold Spring Harbor Laboratory; Chair, Simons Center for Quantitative Biology: *Reconstructing Ancient Human History from DNA*.

July 25: Douglas Fearon, M.D., Professor, Cold Spring Harbor Laboratory; **Robert Maki, M.D., Ph.D.**, Professor, Hofstra Northwell School of Medicine and Professor, CSHL: *Immunotherapy & Cancer—The Latest Research*; cosponsored by CSHL, US Trust, Northwell Health, and St. Johnland Nursing Center.

August 23: Michael Ronemus, Ph.D., Research Assistant Professor, Cold Spring Harbor Laboratory: *Cocktails & Chromosomes*.

October 19: David Jackson, Ph.D., Professor, CSHL; **Zachary Lippman, Ph.D.**, Professor, CSHL; **Doreen Ware, Ph.D.**, Adjunct Associate Professor, CSHL & USDA Agricultural Research Service: *The Changing Relationship between Humans and Plants—"It's Complicated."*

October 25: Gurinder "Mickey" Atwal, Ph.D., Associate Professor, Cold Spring Harbor Laboratory: *Cocktails & Chromosomes*.

November 5: Jonathan Weiner, Pulitzer prize–winning author and Maxwell M. Geffen, Professor of Medical and Scientific Journalism, Columbia Journalism School: *LONG FOR THIS WORLD—Writing about Immortality … and Other Controversial Topics in the Science of Life*; 2017 Lorraine Grace lectureship on societal issues of biomedical research.

CSHL Public Concerts

April 21: Gleb Ivanov and Dmitri Berlinsky, piano and violin
April 28: Anna Polonsky and Orion Weiss, piano duo
May 5: Jocelyn Ho, piano
August 25: Verona Quartet, string quartet
September 8: Matthew Graybil, piano
September 15: Tchaikovsky Trio

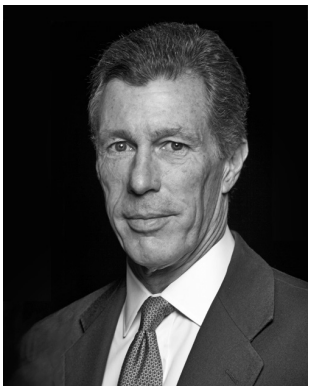
Verona Quartet

Looking Forward

CSHL breakthroughs in research and education will undoubtedly change the world for the better, and I thank all of those who contributed to the Laboratory's mission in 2017.

Bruce Stillman, Ph.D., F.R.S.
President and Chief Executive Officer

CHIEF OPERATING OFFICER'S REPORT



One never tires of reporting positive news and impressive results. The year 2017 was the best in a succession of good years for the Laboratory. Strong financial results from operations and a balanced budget were driven by higher revenues and below-budget expenses. The endowment funds reached a high-water mark as a consequence of successful fund-raising and healthy investment returns. Most impressive, in a tight federal funding environment, the federal grant application success rate of our faculty was nearly double the national average. There is no clearer indication of the high quality of our scientists and the excellence of the research they do every day.

In his President's Letter, Bruce Stillman asks what makes some institutions, like Cold Spring Harbor Laboratory, more productive than others. He cites, among other things, our youthful culture and our commitment to attracting the very best scientific minds regardless of their country of origin. He also mentions a flat organizational structure and an administration that serves the science. Having led the administration now for nearly 20 years, I can confirm that our overriding goal has been to serve and support the academic community and to make this the easiest, the least bureaucratic, and the best place on the planet to do science. This is not a simple task in a rapidly growing organization. Consider the following: Over the last two decades, the Laboratory's annual operating budget has increased from $54 million to $160 million. Total assets have grown from $258 million to $937 million. Endowment funds, net of annual spending, have increased from $160 million to $530 million. Square footage of laboratory research space has increased by 35%. The faculty has grown by 50% to 55 principal investigators. Total head count is up to 1100 from 590.

Cold Spring Harbor has managed to successfully plan for, finance, execute, and absorb this rapid growth. More importantly, we have, in my opinion, added the necessary structure and resources to efficiently operate a growing organization without compromising a highly productive academic culture. This ability to manage growth while preserving the culture and the highest level of overall academic excellence is what distinguishes Cold Spring Harbor Laboratory and makes it a special place to live and work. It takes an extraordinary collection of people and talent to achieve all of this and to carry the culture forward—scientific leaders with uncommon vision, a dedicated and enlightened Board of Trustees, investigators who are driven by the challenge and joy of solving problems, educators and students who are steeped in the scientific process, a committed and expert group of administrative department heads, and a support staff that takes pride in the institutional mission.

As I pen my last annual report and prepare to step down from this wonderful position, I look back on these 20 years with immeasurable satisfaction and pride about what has been accomplished in this amazing place by an unparalleled group of people. As I remarked recently to an academic colleague and friend upon her induction into the National Academy of Sciences, "what a privilege for me to be friends with my heroes and heroines!"

It has been a privilege and an honor, indeed. Thank you all.

W. Dillaway Ayres, Jr.
Chief Operating Officer

Long-Term Service

Front row (*left to right*): Denise Weiss, Nancy Hodson, Inez Sialiano, Sadie Arana, Joseph Houser, Pat Urena, Sal Serafino, Patty Bird, Bob Collins, Robert Eifert, Maria Mosquera, Liz Janow, Kim Bronson, Scilla Wu, Bill Dickerson, Terri Grodzicker, Julie Ehrlich, Aigoul Nourjanova; *back row* (*left to right*): Jan Witkowski, John Inglis, Dave Micklos, Jorge Ramirez, David Spector, Joseph Simorowski, Dick McCombie, Randy Jones, Dill Ayres, Bruce Stillman, Art Brings, Dave Jackson, Bobbie Peters, James Watson.

The following employees celebrated milestone anniversaries in 2017:

45 years	Terri Grodzicker
35 years	Patricia Bird, Arthur Brings, David Micklos
30 years	Sadie Arana, Lisa Bianco, Julie Ehrlich, John Inglis, Barbara Peters, Inez Sialiano, Patricia Urena, Jan Witkowski
25 years	Robert Collins, Nancy Hodson, Randal Jones, W. Richard McCombie, John Pisciotta, Claudia Schmid, Mary Smith, Leslie Wenzel
20 years	Kimberly Bronson, Robert Dickerson, William Dickerson, Robert Eifert, Joseph Houser, David Jackson, Elizabeth Janow, Maria Mosquera, Andrea Newell, Aigoul Nourjanova, Jorge Ramirez, Wilson Ramones, Salvatore Serafino, Joseph Simorowski, Denise Weiss, Scilla Wu

(*Left to right*) Peter Dale, James Watson, Laurie Moller, Gerry Holler, David Spector, Betsy Panagot, Stephanie Goldsmith, Rhonda Veros, Bruce Stillman, Linda Sussman, Graham Wildt, Caizhi Wu, Katie Raftery, Pam Moody, Elizabeth Cherian-Samuel.

15 years	Ingrid Amaya, Cynthia Blaut, Elizabeth Cherian-Samuel, Peter Dale, Stephanie Goldsmith, Gerard Holler, Adam Kepecs, Laurie Moller, Pamela Moody, Elizabeth Panagot, Katherine Raftery, Jose Reyes, Theresa Saia, Linda Sussman, Rhonda Veros, Graham Wildt, Caizhi Wu, Jennifer Ziegler

RESEARCH

CANCER: GENE REGULATION AND CELL PROLIFERATION

Camila dos Santos' laboratory studies the epigenetic regulation of normal and malignant mammary gland development, with an emphasis on the alterations brought by pregnancy. Significant changes mark the pre- and postpubescence mammary developmental stages, but those associated with pregnancy have the greatest effect on cellular function, tissue reorganization, and breast cancer susceptibility. Her group has recently found that mammary glands react differently to a second pregnancy than they do to the first one, with associated changes in DNA methylation. These findings suggested that pregnancy changes the state of mammary cells, and these changes may permanently alter how the cells react to the next pregnancy. In addition, the dos Santos laboratory is exploring how the pregnancy-induced epigenetic changes might influence cell transformation and the risk of breast cancer. This research uses genomic and computational approaches to define the pre- and postpregnancy mammary epigenome. An additional objective of the dos Santos laboratory is to use functional genomics to discover novel transcriptional regulators that modulate mammary stem cell self-renewal, lineage specification, and cell transformation. The long-term objective of Camila's group is to improve the notion of the mammary epigenome during normal development and use this information to gain insight into new preventive and curative strategies to target breast cancer.

Human development requires the regulated activity of thousands of genes in hundreds of distinct cell types throughout life. One requirement for this process is that each cell must contain an intact, functional genome free from mutations. One type of mutation can arise from the activation of transposable elements (TEs). These viral-like parasites lay dormant within our genomes, but they have the capacity to hop into new genomic locations, causing mutations as they break the surrounding DNA sequence. Mounting evidence has implicated transposon activity in a host of human diseases, with particular evidence for TE activation in neurodegenerative diseases: amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD). This is the focus of the research in **Molly Hammell**'s laboratory.

In **Leemor Joshua-Tor**'s lab, researchers study the molecular basis of nucleic acid regulatory processes using the tools of structural biology and biochemistry. One such regulatory process is RNA interference (RNAi), in which a small double-stranded RNA triggers gene silencing. Joshua-Tor and her team offered critical insight when they solved the crystal structure of the Argonaute protein and identified it as the long-sought Slicer. They then went on to explore the mechanism of the slicing event. The structure of human Argonaute 2 (hAgo2) bound to a microRNA (miRNA) guide allowed Joshua-Tor and her colleagues to understand how mRNA is cleaved during RNAi. This year, members of the Joshua-Tor lab explored the function of a very similar protein, called Argonaute 1, that has no slicing ability, although it is almost identical in structure to the slicing hAgo2. Using biochemical methods and mutational analysis, they were able to identify key parts of the protein that are required for slicing activity. The lab also studies the generation of PIWI-interacting RNAs (piRNAs), which serve to protect the genome of germ cells. With colleagues in the Hannon lab, Joshua-Tor's team also determined the structure and function of Zucchini, a key nuclease in the initial generation of piRNAs in fruit flies. In other work, the lab is exploring the mechanisms of heterochromatin formation and gene silencing through the study of a protein complex called RNA-induced initiation of transcriptional gene silencing (RITS). Joshua-Tor is also well known for her work on the E1 helicase enzyme, which acts to unwind DNA strands during the DNA replication process.

Adrian Krainer's lab studies the mechanisms of RNA splicing, ways in which they go awry in disease, and the means by which faulty splicing can be corrected. In particular, they study splicing in spinal muscular atrophy (SMA), a neuromuscular disease that is the leading genetic cause of death in infants. In SMA, a gene called *SMN2* is spliced incorrectly, making it only partially functional. The Krainer lab found a way to correct this defect using a powerful therapeutic approach. It is possible to stimulate SMN protein production by altering mRNA splicing through the introduction into cells of chemically modified pieces of RNA called antisense oligonucleotides (ASOs). Following extensive work with ASOs in mouse models of SMA, one such molecule, known as nusinersen or Spinraza™, was taken to the clinic, and at the end of 2016 it became the first U.S. Food and Drug Administration (FDA)-approved drug to treat SMA, by injection into the fluid surrounding the spinal cord. The Krainer lab is currently using this approach for the study of other diseases caused by splicing defects, including familial dysautonomia. In addition, they are applying antisense technology to stabilize mRNAs that are destroyed by a process called nonsense-mediated mRNA decay (NMD), both to learn about the underlying mechanisms and to develop new therapies. The Krainer lab has also worked to shed light on the role of splicing proteins in cancer. They found that the splicing factor SRSF1 functions as an oncogene, and they recently characterized the splicing changes it elicits when overexpressed in the context of breast cancer; several of these changes contribute to various aspects of cancer progression. Finally, the lab continues to study fundamental mechanisms of splicing and its regulation, and they identified novel ways in which the U1 snRNA can recognize natural 5′ splice sites that deviate from the consensus.

David L. Spector's laboratory is focused on characterizing long noncoding RNAs (lncRNAs) that show altered levels of expression in breast cancer progression and during embryonic stem cell differentiation. A major focus of their efforts has been on Malat1 lncRNA, which is one of the most abundant lncRNAs. The Spector lab previously identified a novel mechanism of 3′-end processing of this RNA. More recent studies have revealed that increased levels of Malat1 lncRNA impact breast cancer progression and metastasis. Knockout or ASO knockdown of Malat1 results in the differentiation of mammary tumors and a significant reduction in metastasis. Studies are currently under way to elucidate the mechanism of action of this abundant nuclear-retained lncRNA and implement innovative therapeutic approaches that can impact its function in vivo. Further, they have identified additional lncRNAs, termed mammary tumor-associated RNAs, that are upregulated in breast tumors, and they are currently assessing the function of these lncRNAs using 3D tumor organoids as well as mouse models.

A second area of study in the Spector lab is based on their earlier discovery of an increase in random autosomal monoallelic gene expression on the differentiation of mouse embryonic stem cells to neural progenitor cells. These data support a model in which stochastic gene regulation during differentiation results in monoallelic gene expression, and for some genes, the cell is able to compensate transcriptionally to maintain the required transcriptional output of these genes. Therefore, random monoallelic gene expression exemplifies the stochastic and plastic nature of gene expression in single cells. Ongoing studies are examining the relationship of monoallelic gene expression to lineage commitment.

Arne Stenlund and colleagues have obtained a detailed understanding of processes required for initiation of DNA replication from the papillomavirus, using this system to gain a general biochemical understanding applicable in other systems. Papillomaviruses are a large viral family that induces cell proliferation at the site of infection, usually giving rise to benign tumors. But certain types of human papillomaviruses (HPVs) generate tumors that progress toward malignancy. Among these are HPVs that cause most cervical cancers. Members of the Stenlund lab also pursue studies aimed at developing an effective small-molecule inhibitor of HPVs that might someday be

used by women who do not receive the preventive anti-HPV vaccine now available or those already infected with HPV who would not be helped by the vaccine.

Bruce Stillman's lab studies the process by which DNA is copied within cells before they divide in two. Working with yeast and human cells, Stillman and colleagues have identified many of the cellular proteins that function at the DNA replication fork during the S phase, the portion of the cell-division cycle when DNA synthesis occurs. Among these proteins are those that facilitate the assembly of chromatin, the protein–DNA complexes that form the chromosomes. Current research focuses on the mechanism that initiates the entire process of DNA replication in eukaryotic cells. At the heart of this mechanism is a protein that binds to "start" sites on the chromosomes, called the origin recognition complex, ORC. The Stillman lab is part of an ongoing collaboration that determined the cryo-EM structure of ORC proteins in complex with a group of proteins, called a helicase, that unwind DNA during replication. These images offer molecular insights into how the helicase is loaded onto DNA. Stillman's research team also focuses on the process by which duplicated chromosomes are segregated during mitosis. They found ORC at centrosomes and centromeres, structures that orchestrate chromosome separation during mitosis. At centromeres, ORC subunits monitor the attachment of duplicated chromosomes to the mitotic spindle that pulls the chromosomes apart when they are correctly aligned. Stillman's team has discovered that mutations in the Orc1 protein alter the ability of ORC to regulate both DNA replication and centrosome duplication. These mutations have been linked to Meier–Gorlin syndrome, a condition that results in people with extreme dwarfism and small brain size, but normal intelligence.

Cancer can be understood as a disease of dysfunctional gene expression control. Research in **Chris Vakoc**'s lab investigates how transcription factors and chromatin regulators cooperate to control gene expression and maintain the cancer cell state. This work makes extensive use of genetic screens to reveal cancer-specific functions for transcriptional regulators, as well as genomic and biochemical approaches to identify molecular mechanisms. One theme that has emerged from their efforts is that blood cancers are often vulnerable to targeting transcriptional coactivators such as BRD4 and the SWI/SNF chromatin remodeling complex. Vakoc's team showed that chemical inhibition of BRD4 displays therapeutic effects in mouse models of leukemia, a finding that has motivated ongoing clinical trials in human leukemia patients. The Vakoc lab has also developed a CRISPR-Cas9 screening approach that can reveal individual protein domains that sustain cancer cells. Their lab is now deploying this technology in a diverse array of human cancers to reveal therapeutic opportunities and basic mechanisms of cancer gene control.

PROBING THE EPIGENOME TO TARGET BREAST CANCER

C. dos Santos C. Chen M. Feigman S.T. Yang
S. Cyrill M. Moss

To Define a Role for BPTF Inhibition in Blocking Breast Cancer Progression

There is growing emphasis on applying epigenetic-based therapies to target the regulation of carcinogenesis. Our laboratory has established that the bromodomain PHD finger transcription factor (BPTF) is a novel epigenetic regulator that maintains normal mammary stem cell self-renewal. Interestingly, we also found that gene networks controlled by BPTF are often up-regulated in breast cancer cells, suggesting a role for BPTF in breast oncogenesis. In fact, we found that BPTF protein levels are increased in human breast cancer tissue, compared with almost undetectable levels in normal epithelial tissue, suggesting an association between BPTF levels and oncogenesis. During this past year, we focused on defining a role for BPTF in controlling growth and survival of patient-derived breast cancer organoid cultures. To date, we have received approximately 10 luminal-like breast cancer specimens and three triple-negative breast cancer specimens and have successfully established luminal-like and triple-negative breast cancer organoid cultures. Cell viability experiments showed that organoid cultures treated with BPTF small molecule inhibitor are less viable than those treated with dimethyl sulfoxide (DMSO)-vehicle control, suggesting a role for BPTF in maintaining cell survival. Follow-up experiments showed that BPTF inhibition drives apoptosis of breast cancer organoid cultures of all subtypes, suggesting that breast cancer, as a disease, depends on BPTF activity to maintain cell survival properties. Global gene expression analysis (RNA-sequencing, or RNA-Seq) after BPTF inhibition showed increased expression of pathways associated with differentiation and shortened cell life span, with concomitant down-regulation of genes regulated by the major oncogene c-MYC, a major oncogene associated with up to 60% of all breast cancer subtypes. We are currently developing reagents and strategies to characterize in vivo the molecular dependencies of BPTF activity to understand how it controls breast cancer development, maintenance, and metastasis.

Investigation of the Epigenetic Modifications Brought About by Pregnancy

C. dos Santos, M. Moss

For nearly 100 years, population studies have consistently and definitively found that a full-term pregnancy early in life dramatically reduces the lifetime incidence of breast cancer. A number of distinct hypotheses have been advanced to explain the effects of parturition on breast cancer risk. Still, there is a profound lack of solid data about the affected cellular populations. Our previous study established that postpregnancy mammary epithelial cells are epigenetically distinct from their prepregnancy counterparts. We showed that transitions through pregnancy stably changed the levels of DNA methylation of mammary epithelial cells, in addition to enhancing their response to the signals of a consecutive pregnancy. To properly characterize the enhancer landscape of mammary epithelial cells after pregnancy, and its influence on gene reactivation or repression, we set out to map the dynamics of several histone marks (H3K27ac, H3K9me3, H3K4me1, and H3K4me3) and gene expression before, during, and after pregnancy. To date, we have found that hierarchical clustering of RNA expression levels implicates a variety of distinct transcription factors in driving gene expression changes at several stages during pregnancy-induced development of the mammary gland. In addition, ChIP-seq analysis of several histone marks indicates that pregnancy augments the activity of distal regulatory regions, also known as enhancers, supporting the notion that the enhancer landscape of mammary epithelial cells is altered by pregnancy cycles. We are currently

investigating the association of transcription factor binding, gene expression regulation, and enhancer landscape activity with the main goal to elucidate changes that influence gene reactivation in the second pregnancy and, most importantly, pinpoint transcriptional/epigenetic alterations underlying pregnancy-induced breast cancer protection.

The Effects of Pregnancy on Reprogramming the Immune Environment of Mammary Epithelial Cells

C. dos Santos, S.T. Yang, M. Feigman, M. Moss

Cancer represents one of the most important challenges of modern medicine and much effort continues to be put toward the development and improvement of new curative therapies. However, the only way to truly eradicate breast cancer is through prevention. Extensive epidemiological data supports an early pregnancy as the most well validated way to reduce breast cancer risk. Yet, a complete understanding of the molecular basis of this phenomenon remains unknown. Our goal is to mechanistically characterize key features in breast cells to develop novel strategies that can prevent breast cancer development. Previously, our lab found that pregnancy induces a dramatic, long-lasting reorganization of the mammary epithelial epigenome. We have found that many of these epigenetic modifications drive expression of genes associated with immune cell communication. Specifically, we discovered CD1d overexpression in breast epithelium after pregnancy. Given that CD1d is a known marker for antigen presentation, we believe that alteration to CD1d levels after pregnancy and during malignant transformation may govern cancer susceptibility in mammary epithelial cells. To address the role of CD1d, we have artificially induced CD1d surface expression in a retinol-dependent manner in an immortalized mouse mammary epithelial cell line. Currently, we are investigating the transcription regulation of CD1d in mammary epithelial cells, specifically studying transcription factors, RARα, STAT5, and PPAR-γ, the last two of which have been implicated in the pathology of numerous diseases, including cancer. In determining the role of CD1d in recruitment and maturation of immune cells, we will characterize CD1d knockout (KO) mammary epithelial cells during pregnancy-dependent development. In a complementary approach, we will use our newly developed mouse model of pregnancy-induced breast cancer protection to probe for a role of CD1d KO mammary epithelial cells (MECs) on decreasing mammary malignant development. We propose that uncovering the role of CD1d will advance the understanding of the molecular basis of pregnancy-induced breast cancer protection. Furthermore, we expect that these discoveries will lay the foundation for the development of rationally designed and molecularly targeted preventative strategies that could reduce the incidence of breast cancer, irrespective of a woman's reproductive history.

The Characterization of Epigenetic Regulators Driving Breast Cancer Initiation and Progression

C. dos Santos, S.T. Yang, S. Cyrill

The association between pregnancy and the risk of breast cancer has been described in several studies, with pregnancy playing a protective role against malignant transformation of normal mammary cells. However, little is known about the molecular processes involved and whether they can be exploited for novel therapeutic avenues against breast cancer development. Our lab previously established that pregnancy effectuates lasting epigenetic changes in the murine mammary gland, priming it for a subsequent pregnancy. Comparative transcriptomic analysis between nulliparous (never pregnant) and parous (postpregnancy) murine mammary epithelial cells pointed toward several chromatin remodeling factors being differentially expressed before and after pregnancy. Therefore, our goal is to identify specific chromatin remodeling proteins that are essential to implement pregnancy-induced protection from breast cancer. During the past year, we developed CRISPR-Cas9 genomic editing reagents to target a series of chromatin remodeling proteins, and used these reagents to generate KO cell lines. We are currently mapping the molecular effects of knocking out specific chromatin remodeling factors on the enhancer landscape organization and regulation of gene expression in mammary epithelial cells. Additionally, we are defining the effects of these chromatin remodeling factors on growth, engraftment, and pregnancy-dependent development of mammary epithelial cells by means of mammary gland transplantation assays. The long-term goal of this approach is to translate our

findings into assays that will investigate normal, nulliparous, and parous human breast organoid cultures and their oncogenic potential.

To Define the Epigenomic Landscape of Organ-Specific, Treatment-Resistant Breast Cancer Metastasis

C. dos Santos, S. Cyrill [in collaboration with M. Egeblad, L. Van Aelst, CSHL]

The most aggressive breast cancers (stage IV) are characterized by metastasis. The heterogeneity that develops within the primary and metastatic tumors is largely effected by interactions with the organ-specific microenvironments, making the disease more aggressive and difficult to treat. However, not much is known about the organ-specific dependencies of breast cancer metastasis. In collaboration with the Van Aelst and Egeblad labs at CSHL, we will use murine models of metastatic breast cancer to study the role of the host microenvironment in metastasis, with particular emphasis on epigenetic regulation and immunomodulation of these invasive cancer cells. We are currently generating the required tools to establish nonimmunocompromised murine models of metastatic breast cancer that provide more accurate microenvironments for the simulation of metastatic disease than existing breast cancer models. These models are generated using a heterogeneous population of 4TI cells, engineered with fluorescently labeled nuclei to use a "direct-from-tissue" bead-based isolation strategy that minimizes procedural artifacts. Chromatin accessibility and enhancer landscape information derived from these nuclei will be used to identify key epigenetic regulators that play a role in the tissue-specific invasion and drug resistance shown by these metastatic cancer cells.

PUBLICATION

Frey WD, Chaudhry A, Slepicka PF, Ouellette AM, Kinberger SE, Pomerantz WCK, Hannon GJ, dos Santos CO. 2017. BPTF maintains chromatin accessibility and the self-renewal capacity of mammary gland stem cells. *Stem Cell Reports* **9:** 23–31.

INTEGRATING GENOMIC DATA SETS TO UNDERSTAND GENE REGULATION IN DEVELOPMENT AND DISEASE

M. Hammell Y. Hao, Y-J. Ho, Y. Jin, W-W. Liao, D. Molik, N. Rozhkov, R. Shaw

Human development requires the regulated expression of thousands of genes in hundreds of distinct cell types throughout life. One requirement for this process is that each cell must contain an intact, functional genome. Furthermore, there is a set of rules for how the genes in our genomes coordinate their activities, and these rules are encoded within gene regulatory networks. Many human diseases occur when these basic processes are altered—either through altering the genome itself (as in the mutations seen in cancers) or through altering the way in which genes interact with each other. The focus of the M. Hammell lab is to understand how mutations in our genomes lead to both alterations in the function of the mutated gene itself and the repercussion of these alterations on the hundreds of other neighboring genes within the network. To this end, the Hammell lab uses computational algorithms to integrate multiple types of genomic and transcriptomic sequencing data into models of cellular function. This includes an emphasis on developing novel tools for the statistical analysis of high-throughput data, novel algorithms for modeling the flow of signals through genetic pathways, and, importantly, testing these models using the tools of molecular genetics.

Endogenous Retroviral-Like Elements May Contribute to Neurodegeneration

Y. Jin, N. Rozhkov, R. Shaw

TDP-43 is an RNA-binding protein that is known to control proper processing of many RNA targets in neurons. Mutation of TDP-43 has been associated with a variety of neurodegenerative diseases, including amyotrophic lateral sclerosis (ALS), fronto-temporal lobar degeneration (FTLD), and Alzheimer's disease (AD). However, the normal function of TDP-43 in neuronal development and maintenance has not been fully characterized and few of its mRNA targets have been definitively associated with the neurodegenerative diseases that result from loss of TDP-43 function. In collaboration with the Dubnau lab at Stony Brook University, our group has explored the novel hypothesis that TDP-43 normally plays a large and hitherto uncharacterized role in regulating the expression of transposable elements (TEs). TEs are retroviral-like elements encoded within our genomes whose unregulated expression leads to genetic instability as well as cellular toxicity. Members of the Dubnau lab have shown that TE transcripts are elevated upon expression of mutant, aggregate-prone forms of human TDP-43 in the fly brain and that neurodegeneration results from expression of this hTDP-43 protein. Members of the Hammell lab have shown that TDP-43 binds widely to TE transcripts in mammals, and TDP-43 binding to TEs is lost in human patients diagnosed with FTLD, a disease characterized by TDP-43 proteinopathy. Although these studies support a role for TDP-43 in regulating TE expression, our future goals are centered on determining the role for misregulation of TEs in neurodegenerative disease. Ongoing research in the lab is focused on three main areas: (1) improving our ability to detect active transposons in ALS and FTLD patient samples, (2) examining the degree to which transposons and other retroviral-like elements are interfering with cellular function in ALS and FTLD patient tissues, and (3) understanding the basic biology of how TDP-43 interacts with the general transposon control machinery.

TEs have been historically difficult to study because of their highly repetitive nature. Nearly half of the human genome is composed of TE-derived sequences, with millions of copies of TEs scattered throughout the chromosomes. Although most of

these copies are nonfunctional, thousands of TEs retain the ability to mobilize and create new copies of themselves elsewhere in the genome. The difficulty lies in differentiating these active TEs from the millions of other harmless copies with nearly identical sequences. This presents both a technical challenge for experimentally isolating TE-derived sequences from the genomes of cells and a computational challenge for determining where each TE copy originates in the genome of a particular sample. Members of the lab have recently developed novel statistical inference methods to solve the computational challenge of analyzing TE expression in sequencing studies (Jin and Hammell 2018). These statistical inference methods have been used to examine the basic mechanisms controlling TE expression and activity in germline tissues in collaborative studies (Krug et al. 2017). Ongoing efforts will establish optimized analysis protocols for many different types of TE studies (genome resequencing studies, chromatin association studies, etc.). In addition, several members of the lab are developing optimized protocols for isolating and identifying novel TE insertion sites in the genomes of individual cells, as active TEs provide one source of genetic mutation that can occur somatically in adult cells. Together, these efforts will provide the tools with which to determine the extent of TE activity in patient samples.

SAKE: Analysis Software for Single-Cell RNA-Seq Data Sets

Y-J. Ho, D. Molik

The Cancer Genome Atlas (TCGA) was a large-scale cancer profiling project that aimed to sequence the genetic mutations and expression profiles for hundreds of patients in dozens of different cancers. Rather than finding a few genetic mutations that explain most cancers, these TCGA studies uncovered thousands of mutations with surprisingly little overlap from patient to patient. This makes the task of designing targeted therapies to treat an individual patient's set of cancer-causing mutations a difficult one. Although the genetic mutations showed little pattern of recurrence among patients, the gene expression data did show clear expression patterns, or cancer molecular subtypes. Moreover, expression subtypes are often predictive of patient survival rates, likelihood to metastasize, and response to targeted therapies. For the melanoma cancer samples that our lab generates, members of the lab have also been able to show that these molecular subtypes persist in 2D cultured cell lines, 3D cultured organoid cells growing in matrigel, and 3D tumors grown in a mouse xenograft model and from tumor samples taken directly from patients. However, for all of these samples, the transcriptomes came from bulk tissue or cell populations whose behavior and phenotypes are generally an average over a very large population of tumor cells. This ignores the heterogeneity that exists both among different cell types within a tissue and among different individual cells within a population. This problem is especially important in the context of human cancers, which are continually evolving to develop invasive properties as well as resistance to therapeutics. To address the problems of cellular heterogeneity, the field of biology needs better tools, both in terms of the technology for single-cell sequencing assays and statistical methods for analyzing data from single-cell assays.

Although our technological ability to generate single-cell data sets has improved considerably over the last few years, our ability to analyze these data sets with algorithms that can robustly detect expression patterns in the presence of high noise and sparse sampling is lagging behind. Members of my lab began by thoroughly testing dozens of statistical models for expression clustering to find a method that would be accurate, robust to noise, and computationally efficient for hundreds of libraries. Using thousands of iterative randomizations and a set of gold standard cell types, we have shown that an adapted version of nonnegative matrix factorization (NMF) shows the most robust performance and is computationally efficient enough to rapidly analyze data from hundreds to thousands of cells. Furthermore, this NMF-based clustering method provides quantitative confidence estimates on the number of clusters present, as well as the assignment of any given cell to a cluster, providing a means to assess the accuracy of our methods as compared with other analysis strategies. We have built a user-friendly web browser application to provide this adapted NMF analysis package to the scientific community, which we have entitled SAKE for single-cell analysis and klustering evaluation.

Mechanisms of Acquired Drug Resistance in Melanoma

Y-J. Ho

The genetic basis of melanoma development is fairly well understood, with activating mutations in the oncogene *BRAF* occurring in a majority of melanoma patient tumors, which also harbor hundreds of secondary mutations of unknown impact. Specific inhibitors that target activated *BRAF*, as well as the downstream MAPK/ERK signaling pathway, have been developed, which dramatically reduce the growth of melanoma cells in patients. However, the effects of these drugs typically extend patient life span for 6 months or less, as the tumors rapidly develop resistance to these targeted therapies. Although some tumors resistant to *BRAF* inhibitors acquire additional genetic lesions that elevate MAPK or AKT signaling, most therapy-resistant cell lines establish resistance without a clearly understood mechanism of resistance. Members of the Hammell lab are developing computational algorithms to understand what changes take place in melanoma cells that develop resistance to *BRAF* inhibitor therapy and how these changes relate to the genotype and molecular subtype of different melanoma tumors.

One goal of the lab is to classify melanoma tumors by both their sets of cancer-specific mutations (their genomic subtype) and the gene transcripts they express (their transcriptional or molecular subtype). Large-scale sequencing surveys of melanoma patients have suggested that hundreds of genomic subtypes exist, such that each patient has their own distinct mutational profile, with a few commonly mutated genes alongside thousands of patient-specific mutations. However, using novel algorithms to analyze the sets of gene transcripts expressed by these tumors, we have determined that cutaneous melanoma can be largely classified into a small number of transcriptional subtypes, and these transcriptional subtypes are highly predictive of overall patient survival rates. Moreover, tumors with each distinct subtype can be seen in human patient samples, mouse xenograft models of melanoma, and cell culture systems. This allows for the subtype-specific modeling of melanoma progression in animal and cell culture systems. Ongoing research is using cell culture models of each melanoma subtype to determine whether response to therapy differs for melanoma cells of different subtypes, and whether this can inform the likelihood of developing therapeutic resistance through different cellular pathways.

To this end, we have taken melanoma cells from two very different melanoma subtypes and sequenced hundreds of individual cells as they respond to MAPK inhibitors. This work, in collaboration with N. Anaparthy in Dr. Jim Hicks's lab at the University of Southern California, has allowed us to explore how well these overall patterns we see in bulk cell extracts are recapitulated at the finer-grained single-cell level. We have used the SAKE single-cell analysis package described above to identify patterns in the expression profiles of these single cells and identified several surprising findings. First, most cells do not alter their underlying molecular subtype when they develop resistance to targeted therapies. Second, cells within a given subtype have several options available for pathways to resistance; that is, multiple separate clusters exist within each subset of therapy-resistant cells. Finally, a small minority of naïve cells that have never been exposed to cancer therapeutics are already expressing markers of therapy-resistant cells. This has led us to a model in which a small number of cells with some degree of innate resistance to targeted therapies provides a platform for adaptation on which the cells elaborate and evolve.

PUBLICATIONS

Ho YJ, Anaparthy N, Molik D, Aicher T, Patel A, Hicks J, Hammell M. 2017. SAKE, single-cell RNA-Seq analysis and klustering evaluation, identifies markers of resistance to targeted BRAF inhibitors in melanoma cell populations. bioRxiv doi: 10.1101/239319. *Genome Res* (in revision).

Krug L, Chatterjee N, Borges-Monroy R, Hearn S, Liao WW, Morill K, Prazak L, Chang YH, Keegan R, Rozhkov N, et al. 2017. Retrotransposon activation contributes to neurodegeneration in a *Drosophila* TDP-43 model of ALS. *PLoS Genet* **13:** e1006635.

Zhang S, Fan G, Hao Y, Hammell M, Wilkinson JE, Tonks NK. 2017 Suppression of protein tyrosine phosphatase N23 predisposes to breast tumorigenesis via activation of FYN kinase. *Genes Dev* **31:** 1939–1957.

In Press

Ho YJ, Anaparthy N, Molik D, Mathew G, Aicher T, Patel A, Hicks J, Hammell MG. 2018. Single-cell RNA-seg analysis identifies markers of resistance to targeted BRAF inhibitors in melanoma cell populations. *Genome Res* **28:** 1353–1363.

Jin Y, Hammell M. 2018. Analysis of RNA-Seq data using TE transcripts. *Methods Mol Biol* **1751:** 153–167.

Nicolas A, Kenna KP, Renton AE, Ticozzi N, Faghri F, Chia R, Dominov JA, Kenna BJ, Nalls MA, Keagle P, et al. 2018. Genome-wide analyses identify *KIF5A* as a novel ALS gene. *Neuron* **97:** 1268–1283.

STRUCTURAL BIOLOGY OF NUCLEIC ACID REGULATORY PROCESSES

L. Joshua-Tor B. Bibel, E. Elkayam, A. Epstein, C. Faehnle, J. Gardiner IV, S. Goldsmith, J. Ipsaro, M. Jaremko, J. Jung, K. Meze, K. On, A. Tocilj

We study the molecular basis of nucleic acid regulatory processes—RNA interference (RNAi) and DNA replication, in particular. We use the tools of structural biology, biochemistry, and biophysics to study proteins and protein complexes associated with these processes to elucidate how they work. X-ray crystallography, cryo-electron microscopy, and other structural techniques enable us to obtain the three-dimensional structures of these molecular machines. Biochemistry, biophysics, and molecular biology allow us to study properties that can be correlated to their function and biology.

Mechanisms of RNAi and Noncoding RNAs

RNAi has made an enormous impact on biology in a very short period of time. Not only are we still discovering new cellular pathways for the regulation of gene expression that use these pathways, but RNAi has become an extraordinary, useful, and simple tool for gene silencing. Almost from its beginnings, people have used genetics, biochemistry, molecular biology, and bioinformatics to study the mechanism of RNAi and related pathways. To get a true mechanistic understanding of these pathways, however, we must understand how the components of the RNAi machinery work at a molecular level.

The let-7 Regulatory Network

C. Faehnle, K. Meze, J. Walleshauser

Although many studies focus on the identification of microRNA (miRNA) targets and the various downstream mechanisms of gene silencing, we are interested in the regulation of a particular miRNA, let-7. The pluripotency factor Lin28 inhibits the biogenesis of the let-7 family of mammalian miRNAs. Lin28 is highly expressed in embryonic stem cells and has a fundamental role in regulation of development, glucose metabolism, and tissue regeneration. Alternatively, Lin28 overexpression is correlated with the onset of numerous cancers, whereas let-7, a tumor suppressor, silences several human oncogenes. Lin28 binds to precursor let-7 (pre-let-7) hairpins, triggering the 3′ oligouridylation activity of TUT4/7. The oligoU tail added to pre-let-7 serves as a decay signal, as it is rapidly degraded by the exonuclease Dis3L2. Genetic disruption of *DIS3L2* is the primary cause of Perlman syndrome, a congenital disorder leading to fetal overgrowth and an increased susceptibility to Wilms' tumor development. Subsequent studies have shown that Wilms' tumors, a common pediatric kidney cancer, overexpress Lin28, underscoring the role of miRNA regulation in kidney tumorigenesis. In somatic cells, in the absence of Lin28, TUT4/7 promotes let-7 biogenesis by catalyzing single uridine addition to a subset of pre-let-7 miRNAs. We are studying the molecular basis and mechanism of Lin28-mediated recruitment of TUT4/7 to pre-let-7, and its effect on the uridylation activity of TUT4/7, switching it from a monouridylation activity to an oligouridylase, and the subsequent degradation of pre-let-7 by Dis3L2.

We found that TUT4 and TUT7 use two multidomain functional modules during the switch from monoU to oligoU. The catalytic module (CM) is essential for both activities, whereas the Lin28-interacting module (LIM) is indispensable for oligoU activity (Fig. 1A). We determined the structure of the TUT7 CM in complex with a group II pre-let-7 double-helical stem and UTP nucleotide trapped in the monoU activity state. This structure revealed a duplex-RNA-binding pocket that orients group II pre-let-7 hairpins to favor monoU addition (Fig. 1B,C).

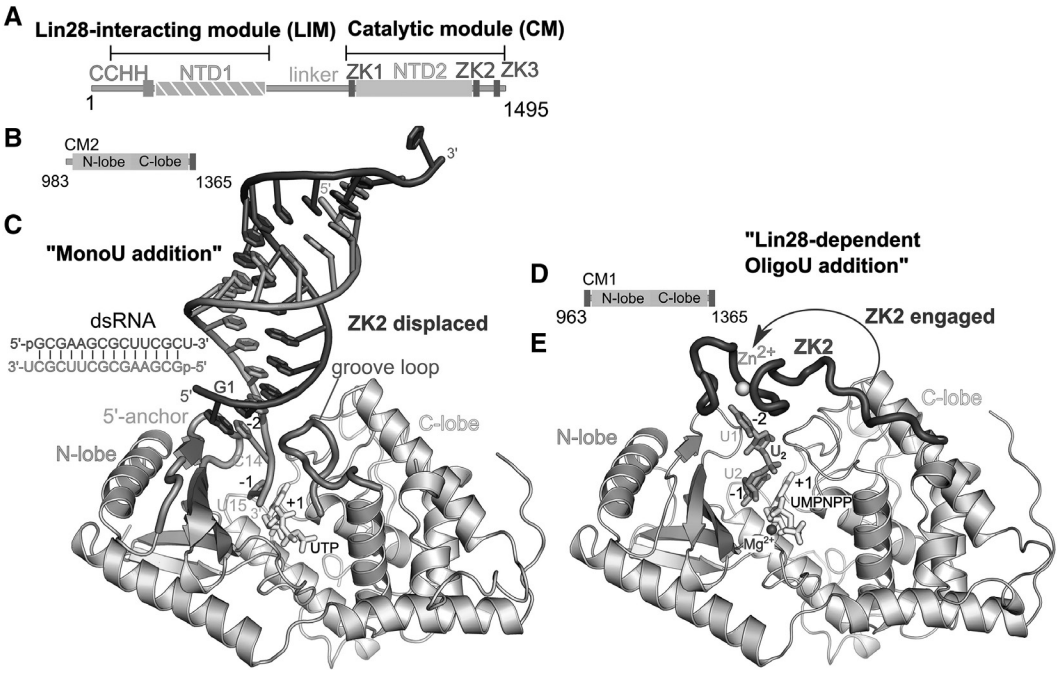


Figure 1. Structures of human TUT7 catalytic module in complex in monouridylation and oligouridylation modes.

The switch to oligoU requires the zinc knuckle (ZK) domain of Lin28 to drive the formation of a stable ternary complex between pre-let-7 and the LIM. To understand the switch from monoU to Lin28-dependent oligoU addition by TUT4/7, we determined structures of the CM in complex with a 2-nt oligoU RNA substrate and a nonhydrolyzable UTP analog, UMPNPP (CM-U2), as well as in complex with a 5-nt oligoU RNA representing the pre- and postcatalytic states of oligo-uridylation. Strikingly, ZK2 of TUT7 snakes between the N- and C-lobes, where it engages the substrate U nucleotide in the –2 position (Fig. 1D,E). In contrast, the nucleotide in the –2 position (C14) in the CM-dsRNA structure is base-paired with G_1, displacing ZK2 (Fig. 1C).

We suggest a model for Lin28-dependent oligoU addition in which ZK2 aids in recycling the enzyme from the post to precatalytic state to allow incoming UTP to access the +1 site for successive rounds of processive uridylation with ZK2 accepting successive U's in the –2 position from the growing oligoU chain. Lending support to this model, both mutant and truncated ZK2 constructs added shorter oligoU tails (~10 U tail vs. >30 U tail) to pre-let-7 in our biochemical assay.

Mechanisms of miRNA-Mediated Silencing

B. Bibel, E. Elkayam

MiRNAs are short noncoding RNAs, 20–24 nt in length, that are loaded into Argonaute proteins to form the RNA-induced silencing complex (RISC). Base-pairing complementation then guides RISC to its target mRNA, and silencing of the target mRNA is achieved by multiple pathways that ultimately lead to degradation of the target mRNA. The physical interaction between human Argonaute and GW182 is essential for facilitating the downstream silencing of the targeted mRNA.

GW182 can interact with Argonaute via three of the GW/WG repeats in its Argonaute-binding domain: motif-1, -2, and the hook motif. We determined a structure of human Argonaute-1 in complex with the hook motif of hGW182 and revealed a "gate"-like interaction, which is critical for GW182 docking into one of Argonaute-1's tryptophan-binding pockets. We also showed that both Argonaute-1 and -2 have a single GW182-binding site. Importantly, miRNA binding increases human Argonaute's affinity to GW182. With target binding occurring rapidly, this ensures that only mature RISC would be recruited for silencing. Using negative-stain single-particle electron

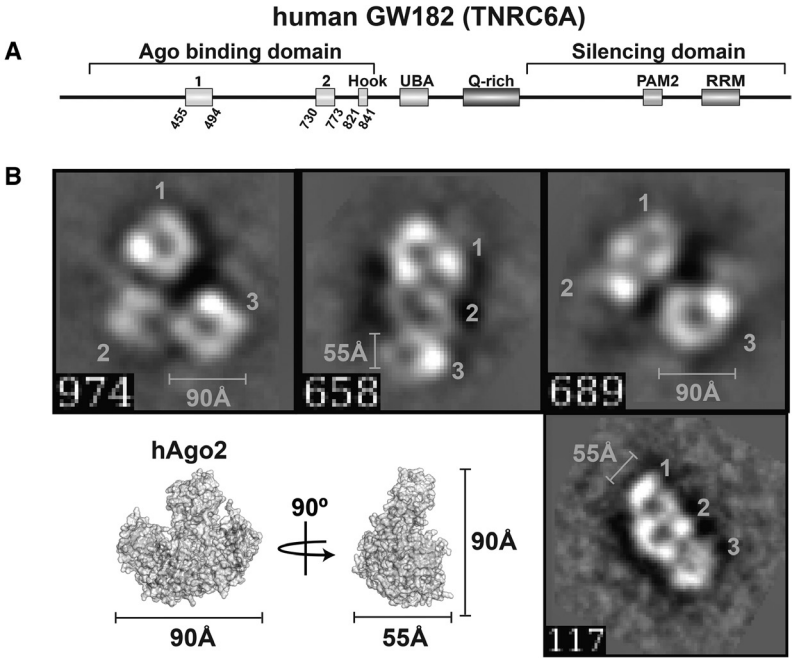


Figure 2. Human GW182 can use all three GW-binding motifs to recruit multiple copies of Argonaute. (*A*) Schematic domain organization of hGW182 (TNRC6A). The different Ago-binding motifs within the Ago-binding domain are indicated as 1, 2, and Hook. (*B*) Selected negative-stain single-particle electron microscopy two-dimensional (2D) averages showing three hAgo2s linked together in the presence of hGW182. The number in the *lower left* corner of each panel refers to the number of raw particles that contributed to the class average. The dimensions of the hAgo2 structure were measured in two different orientations (longest and shortest) and used as a reference to estimate the size of each of the subparticles.

microscopy, we showed that hGW182 can recruit up to three copies of Argonaute via its three CW motifs. This may explain the observed cooperativity in miRNA-mediated gene silencing (Fig. 2).

DNA Replication

DNA replication is the most basic of life processes. Although the structure of Watson and Crick had immediate implications as to how the genetic material is copied, understanding the players involved, the intricate regulation, and even the simple mechanics of this process has been a subject of intense investigation ever since.

The Active Form of the Human Origin Recognition Complex

M. Jaremko, K. On, A. Tocilj [in collaboration with B. Stillman, H. Li, CSHL]

Binding of the origin recognition complex (ORC) to origins of replication marks the first step in the initiation of replication of the genome in all eukaryotic cells. We have been collaborating with Dr. Bruce Stillman here at Cold Spring Harbor Laboratory to understand the molecular mechanism of replication initiation. We determined the structure of the active form of human origin recognition complex (HsORC) by X-ray crystallography and cryo-electron microscopy. The complex is composed of two lobes. The first, composed of the three nominally active ATP-hydrolyzing subunits, ORC1/4/5, which we named the motor module lobe, is organized in an architecture reminiscent of the DNA polymerase clamp loader complexes. A second lobe contains the ORC2/3 subunits. The complex is organized as a double-layered shallow corkscrew, with the AAA+ and AAA+-like domains forming one layer, and the winged-helix domains (WHDs) forming a top layer. CDC6 fits easily between ORC1 and ORC2, completing the ring and the DNA-binding channel, forming an additional ATP hydrolysis site. Analysis of the ATPase activity of the complex provides a basis for understanding ORC activity, as well as molecular defects observed in Meier–Gorlin syndrome mutations.

PUBLICATIONS

Elkayam E, Faehnle CR, Morales M, Sun J, Li H, Joshua-Tor L. 2017. Multivalent recruitment of human Argonaute by GW182. *Mol Cell* **67:** 646–658.

Faehnle CR, Walleshauser J, Joshua-Tor L. 2017. Multi-domain utilization by TUT4 and TUT7 in control of let-7 biogenesis. *Nat Struct Mol Biol* **24:** 658–665.

Ipsaro JJ, Shen C, Arai E, Xu Y, Kinney J B, Joshua-Tor L, Vakoc CR, Shi J. 2017. Rapid generation of drug resistance alleles at endogenous loci using CRISPR-Cas9 indel mutagenesis. *PLoS ONE* **12:** e0172177.

Tocilj A, On KF, Yuan Z, Sun J, Elkayam E, Li H, Stillman B, Joshua-Tor L. 2017. Structure of the active form of human Origin Recognition Complex and its ATPase motor module. *eLife* **6:** e20818.

In Press

Krishnan N, Bonham CA, Rus IA, Kumar Shrestha O, Gauss CM, Haque A, Tocilj A, Joshua-Tor L, Tonks NK. 2018. Harnessing insulin- and leptin-induced oxidation of PTP1B for therapeutic development. *Nat Commun* **9:** 283.

Li J, Wu C, Wang W, He Y, Elkayam E, Joshua-Tor L, Hammond PT. 2018. Structurally modulated co-delivery of siRNA and Argonaute 2 for enhanced RNA interference. *Proc Natl Acad Sci* doi:10.1073/pnas.1719565115.

RNA SPLICING

A.R. Krainer M. Daley, V. Kergourlay, Y.J. Kim, J. Layne, K-T. Lin, Y.H. Liu, W.K. Ma, L. Manche, L. Marasco, J. Mola, F. Papaleonidopoulou, M. Rahman, J. Scharner, D. Segovia, L. Sheng, M.S. Wong, Q. Zhang

Mechanisms of Constitutive and Alternative Pre-mRNA Splicing

RNA splicing is required for expression of most eukaryotic protein-coding genes. The spliceosome selects authentic splice sites with very high fidelity, relying on limited sequence information present throughout introns and exons. In humans, >90% of genes are expressed via alternative splicing, giving rise to multiple protein isoforms. The choice of alternative splice sites is commonly regulated to alter gene expression, either tissue-specifically or in response to a developmental program or signaling pathways. The fact that multiple protein isoforms can be expressed from individual genes shows that the classical "one gene–one enzyme" paradigm is no longer valid and provides an explanation for the unexpectedly small number of protein-coding genes uncovered by genome-sequencing projects.

Both constitutive and alternative splicing mechanisms involve numerous protein components, as well as five noncoding RNA components that are part of small nuclear ribonucleoprotein (snRNP) particles. The work in our lab focuses on the identification and molecular characterization of protein factors and sequence elements that are necessary for the catalysis and fidelity of splicing and/or the regulation of alternative splice-site selection. We are interested in how the spliceosome correctly identifies the exons on pre-mRNA (messenger RNA), and how certain point mutations in either exon or intron sequences cause aberrant splicing, leading to various human genetic diseases. Related areas of interest include the remodeling of messenger ribonucleoprotein (mRNP) architecture as a consequence of splicing, which influences downstream events such as nonsense-mediated mRNA decay (NMD); the various roles of alternative splicing misregulation in cancer; and the development of effective methods to correct defective splicing or modulate alternative splicing, especially in a disease context. A summary of some of our recently published studies is provided below.

Splicing Alterations in Genetic Diseases and Cancer

Pre-mRNA splicing can contribute to the switch of cell identity that occurs in carcinogenesis. We analyzed a large collection of RNA-sequencing (RNA-seq) data sets and found that splicing changes in genes coding for hepatocyte-specific enzymes, such as *AFMID* (arylformamidase, involved in tryptophan metabolism, and hence NAD^+ production) and *KHK* (ketohexokinase), are associated with the survival and relapse of hepatocellular carcinoma (HCC) patients. We found that the switch of *AFMID* isoforms is an early event in HCC development and is associated with driver mutations in *TP53* (p53 tumor suppressor) and *ARID1A* (AT-rich interaction domain 1A, a SWI/SNF family member). Interestingly, this switch of *AFMID* isoforms is human-specific and not detectable in other species, including primates. We further showed that overexpression of the full-length *AFMID* isoform in HepG2 cells results in a higher NAD^+ level, lower DNA-damage response, and slower cell growth. In short, our integrative analysis uncovered a mechanistic link between splicing switches, de novo NAD^+ biosynthesis, driver mutations, and HCC recurrence.

Familial dysautonomia (FD), or Riley–Day syndrome, is a rare inherited neurodegenerative disorder caused by a point mutation in the *IKBKAP* gene, which results in defective splicing of its pre-mRNA. The mutation weakens the 5′ splice site of exon 20, causing this exon to be partially skipped, thereby introducing a premature termination codon (PTC). Although detailed FD pathogenesis mechanisms are not yet clear, correcting the splicing defect in the relevant tissue(s), and thus restoring normal expression levels of the full-length IKAP protein, could be therapeutic.

Using a two-step screen for splice-switching antisense oligonucleotides (ASOs) targeting *IKBKAP* exon 20 or the adjoining intronic regions, we identified a lead ASO that fully restores exon 20 splicing in FD patient fibroblasts. We also characterized the corresponding *cis*-acting regulatory sequences that control splicing of exon 20. When administered into a transgenic FD mouse model, the lead ASO promoted expression of full-length human *IKBKAP* mRNA and IKAP protein levels in several tissues tested, including the central nervous system. These findings provided insights into the mechanisms of *IKBKAP* exon 20 recognition and preclinical proof of concept for an ASO-based targeted therapy for FD.

Spinal muscular atrophy (SMA) is a common, autosomal-recessive motor-neuron degeneration disorder caused by homozygous deletion or mutation of the survival-of-motor-neuron gene, *SMN1*. A closely related *SMN1* paralog, *SMN2*, is present in all patients and differs from *SMN1* by a C to T transition in exon 7 that causes substantial skipping of this exon, such that *SMN2* expresses only low levels of functional, full-length SMN protein. *SMN2* decreases the severity of SMA in a copy number–dependent manner. We previously developed an ASO that efficiently increases the extent of exon 7 inclusion during splicing of *SMN2* transcripts for therapeutic use in SMA. This translational research was performed in collaboration with Ionis Pharmaceuticals. After extensive preclinical testing, multicenter clinical trials (sponsored by Ionis and Biogen) with the ASO compound, nusinersen, were performed. The drug, which is administered to SMA infants and children by lumbar puncture, was well tolerated at all doses tested, and significant dose- and time-dependent improvements in survival, motor function, and achievement of developmental milestones were observed. Nusinersen (Spinraza™) was approved by the U.S. Food and Drug Administration (FDA) at the end of last year for broad use in all SMA types and all ages. This is the first and so far only approved treatment for SMA.

SMN protein is expressed ubiquitously and functions in RNA processing pathways that include trafficking of mRNA and assembly of snRNP complexes. Importantly, SMA severity is correlated with decreased snRNP assembly activity. In particular, the minor spliceosomal snRNPs are affected, and some U12-dependent introns have been reported to be aberrantly spliced in patient cells and animal models. It is likely that aberrant splicing of genes expressed in motor neurons is involved in SMA pathogenesis, but increasing evidence indicates that pathologies also exist in other tissues. We collaborated with Dr. Brage Andresen (University of Southern Denmark, Odense) to comprehensively analyze multiple tissues from an SMA mouse model by RNA-seq. This study showed elevated U12-dependent-intron retention in all tissues examined, and such minor-intron retention was also observed on small interfering RNA (siRNA) knockdown of SMN in HeLa cells. Furthermore, retention of minor introns was reversed by ASO treatment to restore SMN levels in SMA mice, and many downstream transcriptional changes were also reversed. We proposed that missplicing of several minor-intron-containing ion-channel genes may account for disrupted calcium homeostasis in SMA.

We also collaborated with Dr. John Staropoli (Biogen) to characterize gene-expression changes in an induced mouse model of SMA, with or without ASO therapy. This study involved a technique we previously published, TSUNAMI, which uses an ASO to promote exon 7 skipping in *SMN2*-transgenic mice and thus phenocopy severe SMA in a dose-dependent manner, followed by rescue with a different ASO to restore exon 7 inclusion. As above, we found evidence of widespread intron retention, particularly of minor U12 introns, in the spinal cord of mice 30 days after SMA induction, which was then rescued by the therapeutic ASO. Intron retention was concomitant with a strong induction of the p53 pathway and DNA-damage response, manifesting as γ-H2A.X positivity in neurons of the spinal cord and brain. Widespread intron retention and markers of the DNA-damage response were also observed with SMN depletion in human SH-SY5Y neuroblastoma cells and human-induced pluripotent stem cell–derived motor neurons. We also found that retained introns, which tend to be high in GC content, served as substrates for the formation of transcriptional R-loops. We proposed that defects in intron removal in SMA promote DNA damage, in part through the formation of RNA:DNA hybrid structures, leading to motor-neuron death.

The *SMN1* and *SMN2* genes are nearly identical, except for 10 single-nucleotide differences and a 5-nucleotide indel, present in intron 6, exon 7, intron 7, and exon 8. SMA is subdivided into four main types with type I being the most severe. *SMN2* copy number is a key positive modifier of the disease, but it is

not always inversely correlated with clinical severity. We previously reported the c.859G > C variant in *SMN2* exon 7 as a positive modifier in several SMA patients. In collaboration with Dr. Tom Prior (Ohio State University, Columbus) and former postdoc Dr. Yimin Hua (Soochow University, China), we recently identified A-44G in intron 6 as an additional positive disease modifier, present in a group of patients carrying three *SMN2* copies but displaying milder clinical phenotypes than other patients with the same *SMN2* copy number. One of the three *SMN2* copies appears to have been converted from *SMN1*, but, except for the C6T transition in exon 7, no other changes were detected. In the context of minigenes, *SMN1* C6T displayed an ~20% increase in exon 7 inclusion, compared to *SMN2*. Through systematic mutagenesis, we found that the improvement in exon 7 splicing is mainly attributable to the A-44G transition in intron 6. Using RNA-affinity chromatography and mass spectrometry, we further uncovered binding of the RNA-binding protein HuR to the −44 region, where it acts as a splicing repressor. The A-44G change markedly decreases the binding affinity of HuR, resulting in a moderate increase in exon 7 inclusion. This finding is consistent with our previous ASO screen in intron 6, which had revealed the presence of a splicing silencer within this region.

Although motor-neuron degeneration is the hallmark of SMA, accumulating evidence indicates that it is a multisystem disorder, particularly in its severe forms. Several studies delineated structural and functional cardiac abnormalities in SMA patients and mouse models, yet the abnormalities have been primarily attributed to autonomic dysfunction. A collaborative study with the Hua lab and Ionis showed that cardiomyocytes from a severe SMA mouse model undergo G_0/G_1 cell-cycle arrest and enhanced apoptosis during postnatal development. Microarray and real-time reverse transcription polymerase chain reaction (RT-PCR) analyses revealed that a set of genes associated with the cell cycle and apoptosis are dysregulated in newborn pups. Of particular interest, the *Birc5* gene, which encodes Survivin, an essential protein for heart development, was down-regulated as early as presymptomatic postnatal day 0. Cultured cardiomyocytes depleted of SMN recapitulated the gene-expression changes, including down-regulation of Survivin, and abnormal cell-cycle progression, and overexpression of Survivin rescued the cell-cycle defect. Finally, increasing SMN in SMA mice with a therapeutic ASO improved heart pathology and rescued expression of deregulated genes. These data indicate that the cardiac malfunction of the severe SMA mouse model is mainly a cell-autonomous defect, caused by widespread gene deregulation in heart tissue, particularly of *Birc5*, resulting in developmental abnormalities through cell-cycle arrest and apoptosis.

Small-molecule splicing modifiers have been described that target the general splicing machinery and thus have low specificity for individual genes. Several small molecules that correct the splicing deficit of the *SMN2* gene have been identified, some of which are currently being tested in clinical trials for SMA. We collaborated with scientists at Roche (Basel) and the ETH (Zürich) to show that these small molecules directly affect *SMN2* pre-mRNA splicing in vitro and bind to two distinct sites of the *SMN2* pre-mRNA, presumably stabilizing a ribonucleoprotein (RNP) complex that is critical to the specificity of these small molecules for *SMN2* over other genes. In addition to the therapeutic potential of these molecules for the treatment of SMA, this work has wide-ranging implications for understanding how small molecules can interact with specific quaternary RNP structures.

Mechanistic Links between Splicing and NMD

We previously showed that the splicing factor SRSF1 promotes NMD, a quality-control mechanism that degrades mRNAs with PTCs. We showed that transcript-bound SRSF1 increases the binding of the key NMD RNA helicase UPF1 to mRNAs still in (or associated with) the nucleus, bypassing UPF2 recruitment and promoting NMD. SRSF1 promotes NMD when positioned downstream from PTC, resembling the mode of action of exon-junction complex (EJC) and NMD factors. Moreover, we found that splicing and/or EJC deposition increase the effect of SRSF1 on NMD. Finally, we reported that SRSF1 enhances NMD of PTC-containing endogenous transcripts that result from alternative splicing. These findings revealed an alternative mechanism for UPF1 recruitment, uncovering an additional connection between splicing and NMD. SRSF1's role in mRNA biogenesis, from pre-mRNA splicing to mRNA decay, has broad implications for gene-expression regulation and genetic diseases.

PUBLICATIONS

Doktor TK, Hua Y, Andersen HS, Brøner S, Liu YH, Wieckowska A, Dembic M, Bruun GH, Krainer AR, Andresen BS. 2017. RNA-sequencing of a mouse-model of spinal muscular atrophy reveals tissue-wide changes in splicing of U12-dependent introns. *Nucleic Acids Res* **45:** 395–416.

Jangi M, Fleet C, Cullen P, Gupta SV, Mekhoubad S, Chiao E, Allaire N, Bennett CF, Rigo F, Krainer AR, et al. 2017. SMN deficiency in severe models of spinal muscular atrophy causes widespread intron retention and DNA damage. *Proc Natl Acad Sci* **114:** E2347–E2356.

Sivaramakrishnan M, McCarthy KD, Campagne S, Huber S, Meier S, Augustin A, Heckel T, Meistermann H, Hug MN, Birrer P, et al. 2017. Binding to SMN2 pre-mRNA-protein complex elicits specificity for small molecule splicing modifiers. *Nat Commun* **8:** 1476.

Wu X, Wang SH, Sun J, Krainer AR, Hua Y, Prior TW. 2017. A-44G transition in SMN2 intron 6 protects patients with spinal muscular atrophy. *Hum Mol Genet* **26:** 2768–2780.

In Press

Aznarez I, Nomakuchi TT, Tetenbaum-Novatt J, Rahman MA, Fregoso O, Rees H, Krainer AR. 2018. Mechanism of nonsense-mediated mRNA decay stimulation by splicing factor SRSF1. *Cell Rep* **23:** 2186–2198.

Lin KT, Ma WK, Scharner J, Liu YR, Krainer AR. 2018. A human-specific switch of alternatively spliced AFMID isoforms contributes to TP53 mutations and tumor recurrence in hepatocellular carcinoma. *Genome Res* doi: 10.1101/gr.227181.117.

Sheng L, Wan B, Feng P, Sun J, Rigo F, Bennett CF, Akerman M, Krainer AR, Hua Y. 2018. Downregulation of Survivin contributes to cell-cycle arrest during postnatal cardiac development in a severe spinal muscular atrophy mouse model. *Hum Mol Genet* **27:** 486–498.

Sinha R, Kim YJ, Nomakuchi T, Sahashi K, Hua Y, Rigo F, Bennett CF, Krainer AR. 2018. Antisense oligonucleotides correct the familial dysautonomia splicing defect in IKBKAP transgenic mice. *Nucleic Acids Res* **46:** 4833–4844.

REGULATION OF GENE EXPRESSION

D.L. Spector G. Arun, B. Balasooriya, S. Barrera, B. Benz, K-C. Chang, S. Diermeier, N. El-amine, R. Hazra, B. Liu, S. Russo, W. Xu, A. Yu

Most cellular processes can trace their beginnings to the nucleus where a gene is activated, resulting in the production of an RNA molecule—some of which encode proteins, whereas others are functional in the form of RNA. Although much biochemical information is available regarding many of the factors involved in gene expression, the spatial and temporal parameters that influence gene expression and the role of noncoding RNAs in regulating this multifaceted process are just beginning to be elucidated. Over the past year, our research has continued to focus on identifying and characterizing the role of long noncoding RNAs (lncRNAs) in cancer progression and/or differentiation. In addition, we have been examining the role of lineage commitment in establishing random autosomal monoallelic gene expression and the role of lncRNAs in lineage commitment. Following is an overview of some of our accomplishments over the past year.

Identification of lncRNAs Involved in Breast Cancer Progression

G. Arun, S. Diermeier, K-C. Chang, S. Russo, B. Benz, N. El-amine, B. Liu, W. Xu [in collaboration with C. Frank Bennett, R. MacLeod, F. Rigo, Ionis Pharmaceuticals]

Genome-wide studies have identified thousands of lncRNAs lacking protein-coding capacity. *MALAT1* (metastasis-associated lung adenocarcinoma transcript 1) is among the most abundant and highly conserved nuclear restricted lncRNAs, and its expression is misregulated in many cancers, including breast cancer. Our previous studies have shown that loss of *Malat1* in the *MMTV-PyMT* mouse model of luminal B breast cancer resulted in differentiation of primary tumors and a significant reduction in tumor growth rate. Most interestingly, a marked decrease was observed in the incidence of lung metastases; ~70% fewer metastatic nodules in *Malat1* antisense oligonucleotide (ASO)-treated animals versus scrambled ASO-treated animals. RNA-sequencing (RNA-Seq) analysis of the primary tumors and tumor-derived organoids treated with *Malat1* ASO showed up-regulation of genes involved in differentiation and down-regulation of genes involved in migration and proliferation. Further, *Malat1* knockdown also resulted in altered pre-mRNA splicing of many genes, including critical transcription factors. Over the past year, we showed that loss of *MALAT1/Malat1* can impact preexisting metastases using a breast cancer xenograft model. Human MDA-MB-231-LM2 breast cancer cells harboring a luciferase reporter were intravenously introduced into nude mice. ASO targeting *MALAT1/Malat1* or ScASO was administered after detection of stable luminescence in the lungs of the animals. Interestingly, *MALAT1/Malat1* ASO-treated mice showed a significant decrease in luminescence, or stabilization of the signal, after eight weeks of treatment, whereas the disease burden in control animals increased severalfold as quantified by measuring the photon flux intensity. Further, we have performed RNA-FISH on human breast tumor sections representing various stages, as well as multiple subtypes, of human breast cancer. Higher *MALAT1* expression was observed within subsets of cells in the primary tumor. Such cells or cell clusters increased with severity of the disease and its metastatic state, raising the possibility that these *MALAT1*-high cells represent a unique population of cells that have a higher propensity to metastasize. Together, our data indicate that *MALAT1/Malat1* lncRNA regulates critical processes in breast cancer pathogenesis and represents a promising therapeutic target for metastatic disease.

We previously identified a group of lncRNAs that are up-regulated in mammary tumors versus normal mammary epithelial cells. *Mammary Tumor-Associated RNA 25* (*MaTAR25*) is a nuclear-enriched and chromatin-associated lncRNA with a length of ~2000 nucleotides. *MaTAR25* is overexpressed in mammary

tumors in both the MMTV-PyMT (luminal B subtype) and the MMTV-NEU-NDL (HER2 subtype) mouse models compared with normal mammary epithelial cells. By using the CRISPR-Cas9 genome-editing system, we generated *MaTAR25* knockout (KO) clones in 4T1 cells. When comparing *MaTAR25* KO cells to control 4T1 cells, we found a significant decrease in cell proliferation (−50%), cell motility (−40%), and invasion (−45%). Ectopic expression of *MaTAR25* in 4T1 *MaTAR25* KO cells rescues the cell proliferation and invasion phenotypes, indicating that *MaTAR25* lncRNA plays an important role in these processes. Injection of *MaTAR25* 4T1 KO cells into the mammary fat pad or tail vein of BALB/c mice resulted in a decrease in both tumor growth (−56%) and the number of lung metastatic nodules (−62%) compared with the 4T1 control group. Furthermore, ASO-mediated knockdown of *MaTAR25* in the MMTV-Neu-NDL mouse model resulted in a decrease in tumor growth (−59%) compared with a scrambled ASO control group. Comparison of differentially expressed genes from *MaTAR25* KO cells with *MaTAR25*-targeted genes identified in wild-type cells by chromatin isolation by RNA purification (ChIRP) data sets identified a total of 446 overlapping genes, which are likely downstream targets regulated by *MaTAR25*. Analysis of synteny between the mouse and human genomes, combined with TCGA RNA-Seq data from individuals with breast cancer, has allowed us to identify the human ortholog of *MaTAR25* (*hMaTAR25*), which has been confirmed by rescue experiments. Together, our data suggest that *hMaTAR25* may have diagnostic and/or therapeutic potential as a biomarker and/or a novel drug target in breast cancer progression.

Human Breast Tumor Organoid Project

G. Arun, S. Barrera, S. Russo, K. Kostroff (Northwell Health)

Tumor organoids grown in three-dimensional (3D) culture provide a very innovative and unique platform to study cancer, as they can recapitulate many aspects of the disease with high fidelity. As such, they represent an excellent system for identifying new therapeutic targets and for drug development and screening in a patient-specific manner.

We have established a living Human Breast Tumor Organoid Biobank in collaboration with surgeons and pathologists at Northwell Health. As part of this, we have generated tumor organoids from more than 30 patient tumors spanning a wide spectrum of pathologically distinct breast cancer (BC) subtypes that are present in the population. A significant number of the patient tumor organoids have a matched normal breast tissue organoid from a region that is unaffected by cancer. Ten percent of these 3D organoid models already derived in the lab are from patients with triple-negative breast cancer (TNBC). These tumor organoids will be used to evaluate the impact of *MALAT1* knockdown and KO on the growth properties of these patient-derived organoids. In a pilot experiment, we have successfully knocked down *MALAT1* using ASOs in these patient tumor–derived organoids. Further, imaging-based analysis has shown that *MALAT1* ASO treatment could induce a more differentiated morphology to a subset of the patient tumor–derived organoids. These studies will be expanded to a broad group of patient-derived organoids, and downstream molecular targets will be identified.

The Role of the lncRNA, *Platr4*, in Lineage Commitment

R. Hazra

lncRNAs are differentially expressed in a development-specific manner across tissues, suggesting regulatory roles in cell fate decision and differentiation. We have identified the functional role of a novel embryonic stem cell (ESC)-specific lncRNA, *Platr4* (pluripotency-associated transcript 4) in cell fate determination. *Platr4* is a 986-nucleotide poly(A)$^+$ transcript comprising two exons. Cellular fractionation of mouse ESCs indicates that *Platr4* is mainly present in the nuclear fraction and associated with chromatin. Using the CRISPR-Cas9 genome-editing system, we have generated mouse embryonic stem cell (mESC) lines (V6.5 and AB2.2) with deletion of the *Platr4* promoter, resulting in a significant depletion (homozygous deletion, *Platr4*$^{-/-}$) of the *Platr4* transcript (up to 99%) as measured by qRT-PCR and single-molecule RNA-FISH analysis. Deletion of *Platr4* in mESCs did not affect cell cycle kinetics or pluripotency. In contrast, we identified abnormalities in the spontaneous contraction of embryoid bodies (EBs) in *Platr4*$^{-/-}$ cells, compared with control cells, during the differentiation process. In

control cells, 27% of EBs beat at day 12 compared with 2% in *Platr4*$^{-/-}$ cells. Consistent with these data, the expression of cardiac Troponin T (*cTnT*), an integral component of the contraction machinery, shows decreased levels in *Platr4*$^{-/-}$ EBs. Further, morphological abnormalities of EBs were observed with smaller size and a darker cavity in *Platr4*$^{-/-}$ cells. Moreover, the relative expression levels of transcripts expressed in endoderm (*Sox17*, *Foxa2*) and mesoderm (*Tbx5*, *Gata4*) were markedly reduced in *Platr4*$^{-/-}$ cells compared with control cells. In addition, *Platr4*$^{-/-}$ mESCs showed a significantly reduced expression of the cardiovascular gene network during targeted differentiation of cardiomyocytes. Therefore, our findings indicate that *Platr4* is an important lncRNA regulator of lineage commitment during mammalian development.

Probing the Role of a Highly Expressed lncRNA in Hepatocellular Carcinoma

A. Yu [in collaboration with C. Berasain, University of Navarra, Pamplona, Spain; F. Rigo, Ionis Pharmaceuticals]

Hepatocellular carcinoma (HCC), the most common type of liver malignancy, is one of the most lethal forms of cancer. HCC is not diagnosed until late stages and has a poor 5-year survival rate of <14%. Excluding liver transplantation, the current standard of care for HCC is treatment with sorafenib, a multikinase inhibitor that targets Raf, receptor tyrosine kinases, and platelet-derived growth factor receptor, which extends median survival time from 7.9 months to 10.7 months. This modest gain emphasizes the urgent need to identify new and effective therapeutic targets for HCC. We have determined lnc05 to be highly expressed in mouse HCC cell lines and conserved in human cell lines. In preliminary studies, lnc05 is up-regulated in HCC cells compared with normal mouse hepatocytes in cell lines, as well as during hepatocarcinogenesis. KO of lnc05 in HepA1-6 cells has a 23% increased doubling time or a 47% reduction in proliferation. In addition, weighted-gene correlation analysis (WGCNA) places lnc05 in a module that is enriched with cell-cycle components, and the KO results in 26% increase of cells in S phase and a 41% decrease in G_2. Moreover, knockdown of lnc05 expression using ASOs decreases proliferation, as assessed by expression levels of Ki67, a proliferation marker, and colony formation assays. Ongoing studies are examining the role of lnc05 in HCC.

Random Autosomal Monoallelic Gene Expression and Differentiation

B. Balasooriya [in collaboration with S. Ballouz, J. Gillis, CSHL]

Monoallelic gene expression describes the transcription from only one of two homologous alleles of a particular gene in a diploid cell. Over the past year, we initiated studies to determine the impact of various differentiation paradigms (ectoderm, mesoderm, endoderm) on the establishment of stochastic autosomal monoallelic gene expression at single-cell resolution. We generated paired-end single-cell RNA-seq data from full-length cDNA synthesized using Switching Mechanism at 5′ End of RNA Template and locked nucleic acid technology (SMARTer-seq). We used diploid F_1 hybrid (C57BL/6J-CAST/EiJ) mouse embryonic stem cells (F_1 mESCs) cultured in 2i and leukemia inhibitory factor (LIF). These cells were subsequently differentiated in vitro into cardiac progenitor cells (CPCs) and cardiac stem cell cultures (CSCs). Further, we differentiated CPC- and CSC-containing cultures into mature and functional cardiac cell species (cardiomyocytes, CM; cardiac fibroblast, CFC; vascular smooth muscle cells, vSMC; and endocardial cells, ECC). We first analyzed the single-cell RNA-Seq data using publicly available allele-specific expression analysis pipelines. This single-cell RNA-Seq analysis showed that, in the cardiac lineage, the percentage of random autosomal monoallelically expressed (RAME) genes in CPCs was similar to the percentage of RAME genes detected in F_1 mESC (0.74% ± 0.04%). Although the specific differentially expressed genes between F_1 mESC and CPCs changed, the total number of detected RAME genes in F_1 mESCs and CPCs was not statistically different (ESCs = 21, CPCs = 24). However, in F_1 mESC–derived CFCs, the percentage of RAME genes observed versus the total number of detected genes expressed was ~10-fold less than in F_1 mESCs. We also observed that the total number of expressed genes was ~50% less in CFCs compared with F_1 mESCs, probably as a result of the initiation of a differentiation gene expression program. To confirm and extend these observations and investigate RAME gene expression in the

other cardiac lineage-specific cells with more stringent statistical analytical methods, we are designing a new pipeline in collaboration with Jesse Gillis and Sarah Ballouz. The new pipeline will also address the "evolution of RAME gene expression" throughout the differentiation process in different cell types of the cardiac lineage.

In addition, we are in the process of developing a genetic system to visualize allele-specific gene transcription in living cells.

PUBLICATIONS

Diermeier SD, Spector DL. 2017. Antisense oligonucleotide-mediated knockdown in mammary tumor organoids. *Bio-protocol* **7:** e2511.

Li MA, Amaral PP, Cheung P, Bergmann JH, Kinoshita M, Kalkan T, Ralser M, Robson S, von Meyenn F, Paramor M, et al. 2017. A lncRNA fine tunes the dynamics of a cell state transition involving *Lin28*, *let-7* and *de novo* DNA methylation. *eLife* **6:** e23468.

Zhang B, Mao YS, Diermeier SD, Novikova IV, Nawrocki EP, Jones TA, Lazar Z, Tung C-S, Luo W, Eddy SR, et al. 2017. Identification and characterization of a class of *MALAT1*-like genomic loci. *Cell Rep* **19:** 1723–1738.

MOLECULAR BIOLOGY OF PAPILLOMAVIRUSES

A. Stenlund M. Kornaj

The papillomaviruses are a group of viruses that infect and transform the basal epithelium, inducing proliferation of the cells at the site of infection. The resulting tumors (warts) are, in most cases, benign and will usually regress after some time, but certain types of human papillomaviruses (HPVs) give rise to tumors that are prone to progress toward malignancy, especially frequently cervical carcinoma. Indeed, HPV-6 infection appears to be a necessary cause of invasive cervical carcinoma and, thus, represents one of the few firmly established links between viral infections and the development of cancer.

An impediment to the study of papillomaviruses has been the inability to define simple in vitro cell culture systems for analysis of the viral life cycle. These viruses normally require specialized differentiating cells that, only with difficulty, can be generated in cell culture. However, for a bovine papillomavirus (BPV-1), a convenient cell culture system exists in which viral gene expression, oncogenic transformation, and viral DNA replication can be studied. Thus, BPV has become a useful model for these aspects of the viral life cycle. The DNA replication properties of the papillomaviruses show some unique and interesting characteristics. As part of their normal life cycle, these viruses can exist in a state of latency, which is characterized by maintenance of the viral DNA as a multicopy plasmid in infected cells. The copy number of the viral DNA is tightly controlled, and the viral DNA is stably inherited under these conditions. Papillomaviruses, therefore, provide a unique opportunity to study plasmid replication in mammalian cells. In addition, the viral DNA replication machinery represents one of the most promising targets for antiviral therapy.

In previous years, we have reported the characterization of the papillomavirus replicon and identification of the viral components that are required for viral DNA replication. In recent years, we have directed our attention toward the biochemical events that are associated with initiation of DNA replication. We are studying the biochemical properties of the viral E1 and E2 proteins, and how these two proteins interact with the viral origin of DNA replication and the cellular replication machinery to generate initiation complexes. Our studies show that the E1 protein has all the characteristics of an initiator protein, including origin of replication (ori) recognition, DNA-dependent ATPase activity, and DNA helicase activity. The transcription factor E2, whose precise function has remained more elusive, appears to serve largely as a loading factor for E1. Through direct physical interactions with both E1 and the ori, E2 provides sequence specificity for the formation of the initiation complex.

We are currently attempting to elucidate how the E1 and E2 proteins orchestrate the precise biochemical events that precede initiation of DNA replication at the viral ori. These events include binding of the initiator to the ori, the initial opening of the DNA duplex (melting), as well as the assembly and loading of the E1 replicative helicase at the replication fork. Our studies, so far, indicate that these activities are generated in an ordered process that involves the sequential assembly of E1 molecules on the ori. This sequential assembly generates different complexes with different properties that, in turn, recognize ori, destabilize the double helix, and function as the replicative DNA helicase.

The Formation of Different E1-DNA Complexes Involves the Use of Different Combinations and Permutations of Four Different DNA-Binding Activities Present in E1

The papillomavirus E1 protein is a representative of a subgroup of hexameric helicases. However, E1 is not a dedicated helicase, but also serves several additional functions in the viral life cycle. E1 takes part in initiation of viral DNA replication in several different ways: The E1 DNA-binding domain (DBD) provides sequence-specific binding of E1 to the viral ori. A particular form of E1, a double trimer (DT), can then generate local melting of the viral ori. Once the DNA is melted locally, an E1 DH is formed on the melted DNA and unwinds the DNA in front of the replication fork.

Obviously, the use of a single polypeptide to provide multiple functions presents many challenges. In addition to accommodating multiple different activities in one polypeptide, mechanisms have to exist that allow switching between the different activities. Furthermore, mechanisms have to exist that prevent the different activities from interfering with each other. For example, in the DT to DH transition, E1 switches from site-specific DNA binding and complex formation dictated by E1 binding sites in the ori, to formation of hexamers dictated by the oligomerization domain and binding to single-stranded DNA (ssDNA). How this feat is accomplished is not understood.

We have previously showed that the E1 protein contains two different DNA-binding activities. One of these is the E1 DBD, which binds to specific sites in the origin of DNA replication and directs binding of E1 as a head-to-head dimer. The second DNA-binding activity consists of a β-hairpin structure in the helicase domain, which directs non-sequence-specific DNA binding. These two elements are required to form the DT and DH complexes. In addition to the DT and DH complexes that form on the origin of DNA replication, E1 can form two sequence-independent complexes on double-stranded DNA. These complexes, a trimer in the presence of ADP and a hexamer that depends on ATP hydrolysis, provide information about the intrinsic ability of E1 to oligomerize, and are thought to relate to the formation of the DT and DH, respectively.

We wanted to determine whether the formation of the trimer and hexamer also relies on the same two DNA-binding elements that are important for formation of the DT and DH. We, therefore, generated mutations in the DBD and β-hairpin in the helicase domain, in residues involved in DNA contacts, and tested these for trimer and hexamer formation. Strikingly, although the β-hairpin was required for trimer formation, the DBD was not. However, the DBD was critically important for hexamer formation, demonstrating that the formation of these complexes relies on different types of DNA-binding activities.

Because the β-hairpin by itself is not sufficient for trimer formation, we wanted to examine E1 for additional DNA-binding activities. Although E1 is well studied, functions have not been assigned to all parts of the protein and especially the amino-terminal, and ~150 residues remain mysterious. A part of this sequence is involved in nuclear import and export, but these sequences account for only a fraction of the amino-terminal domain. Given the parsimony that usually characterizes viral proteins, it is likely that additional functions are present in the amino-terminal domain.

We identified a DNA-binding activity present in the amino-terminal domain by generating fragments from this domain and expressing these as GST fusion proteins in *Escherichia coli*. In this manner, we could pinpoint a nonspecific DNA-binding activity located between residues 70 and 120. Point mutations that disrupted the DNA-binding activity of the $E1_{70\text{-}120}$ peptide resulted in a failure to form the DT complex, consistent with a role for this DNA-binding activity in DT formation.

To clarify which DNA-binding activities in E1 were required for the formation of which E1 complex, we tested mutants in the different DNA-binding peptides for complex formation. We found that a total of four DNA-binding activities, in different combinations, are involved in E1 complex formation. To form the trimer complex requires a combination of the β-hairpin and $E1_{102\text{-}120}$ peptide, whereas formation of the hexamer complex requires a combination of the β-hairpin and a nonspecific DNA-binding activity in E1 DBD. To form the DT complex requires three of these DNA-binding activities: the β-hairpin, 102-120 peptide, and the sequence-specific DNA-binding activity present in the E1 DBD, and all play a role in DT formation. Finally, to form the DH complex requires all four DNA-binding activities.

Our data show that the four DNA-binding activities are not all exposed and accessible at the same time. For example, the DNA-binding activity of the $E1_{102\text{–}120}$ peptide is not detectable in the context of the intact amino-terminal domain. However, successive deletions from the amino terminus gradually increase the DNA-binding activity of this fragment. These results show that the DNA-binding activity present in $E1_{70\text{-}120}$ is hidden or sequestered in the presence of the intact amino terminus. Similarly, the E1 DBD can be inactivated through an interaction with a short peptide in the amino-terminal domain. These two interactions are mutually exclusive, resulting in the exposure of either one or the other of the two DNA-binding activities. The toggling between the two different DNA-binding activities allows E1 to transition from one kind of complex (i.e., the DT) to another kind of complex (i.e., the DH). Although we do not understand the exact mechanism of the switch, it is clear that it is ATP binding and hydrolysis by E1 that causes the toggle to change its position.

DNA REPLICATION AND CHROMATIN INHERITANCE

B. Stillman A. Amendolara K. Bhalla H-C. Chou M. Hossain Y. Hu S. Nandi F. Nasrin Y-J. Sheu M. Tramantano P. Wendel

In 1986, we expanded our research from investigating the replication of virus genomes in human cells to understanding the replication of the genome of eukaryotes, at first focusing on the budding yeast *Saccharomyces cerevisiae* and then on the human genome itself. Research in the early 1980s demonstrated that DNA from eukaryotic viruses or from bacteria that had been injected into activated *Xenopus* eggs could be duplicated in a regulated manner, being copied only once per cell division cycle in synchrony with the endogenous cell chromosomes. These observations suggested that at least in this system regulated DNA replication did not require specific DNA sequences that functioned as origins of DNA replication. In contrast, in the late 1970s, it was shown that DNA sequences derived from the genome of *S. cerevisiae* were known to confer on plasmids the ability to be stably maintained as extrachromosomal plasmids in the yeast cells. These "autonomously replicating sequences" (ARS) were later shown to be origins of DNA replication. By 1992, we had mapped in detail the sequences of one of these origins (*ARS1*) and had identified the origin recognition complex (ORC) that bound to elements within the ARSs and showed that ORC was required for replication of the chromosomal DNA. Later, we identified the yeast Cdc6 protein as an essential initiator component that bound directly to ORC. Since then, many proteins have been identified as being essential for the initiation and replication of eukaryotic DNA, using both the eukaryotic virus SV40 and the budding yeast systems. In 2009, the assembly of the pre–replicative complex (pre-RC), containing an Mcm2-7 double hexamer (DH), was reconstituted with purified proteins by two groups—John Diffley's laboratory and a collaboration between Christian Speck's laboratory and this laboratory. In 2017, John Diffley's and Dirk Remus's laboratories reported the reconstitution of complete replication of naked DNA with 24 purified proteins (containing 55 separate polypeptides), as well as the replication of DNA packaged into nucleosomes. All of these accomplishments represent milestones in our long-standing quest to understand the biochemistry of how our genome is inherited.

During this current year, we have focused on discovering how these initiator proteins work, particularly focusing on the structures of the proteins involved in the initiation of DNA replication. Part of this research was performed in collaboration with Christian Speck and structural biologists Huilin Li and Leemor Joshua-Tor.

The Structure and Function of Initiators in Yeast and Human Cells

The 2009 research showed that the Mcm2-7 DNA helicase components that are loaded onto *S. cerevisiae* origins of DNA replication by ORC, Cdc6, and the Cdt1 protein exist as double hexamers that can move on the double-stranded DNA. The previous research had shown that the double hexamers could move along the DNA and were topologically linked to the DNA helix. This year, employing cryo-electron microscopy, we identified the structure of an Mcm2-7 loading intermediate called the OCCM, containing ORC, Cdc6, and one hexamer of the Mcm2-7 complex encircling the double-stranded origin DNA (*ARS1*). The OCCM, formed in the presence of ATPγS that did not allow ATP hydrolysis, still had the Cdt1 protein bound to the Mcm2-7 hexamer, revealing for the first time the complete structure of the three-domain Cdt1 protein, along with the structure of ORC and Cdc6 bound to DNA. We also determined, using cryo-electron microscopy, the high-resolution structure of the Mcm2-7 DH bound to *ARS1* origin DNA, showing that the DNA double helix passes through the middle of the barrel-shaped Mcm2-7 double hexamers and is kinked twice as it binds to the inner surface of the Mcm2-7 barrels. This structure led to a detailed model (Fig. 1) of how the DNA might be unwound when the Mcm2-7 double hexamer is activated by a number of proteins that either bind to the Mcm2-7 DH or phosphorylate components of the DH or its activating proteins.

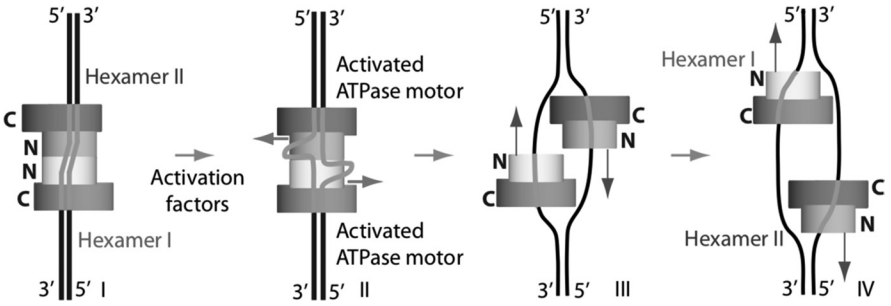


Figure 1. Model for the initial DNA unwinding based on the cryo-EM structure of the Mcm2-7 double hexamer bound to the *ARS1* origin of DNA replication. In the structure, the double-stranded DNA is kinked twice within the two barrel-shaped Mcm2-7 hexamers that form in a head to head configuration (panel *I*). Panel *II* shows the proposed lateral shift and tilt of the *N*-tier rings of the Mcm2-7 hexamers that then separate (panel *III*) and extrude what will become the lagging-strand templates of the replication forks. Panel *IV* shows that the two helicase hexamers would pass each other on opposite strands of the DNA to create two active DNA replication forks. (Reproduced from Noguchi et al. 2017.)

The key finding was that the inner surface of each hexamer had the future lagging strand intimately associated with the junction between the Mcm2 and Mcm5 subunits, the gap through which the dsDNA is loaded in the first place. Thus, one strand of the DNA is associated with one hexamer and the other strand is associated with the second hexamer in the DH structure, providing asymmetry for DNA unwinding.

In addition to the structure of the OCCM and the Mcm2-7 DH using cryo-EM, a collaboration with Leemor Joshua-Tor has revealed the detailed structure of the human ORC, lacking ORC6 and the amino terminus of ORC1. This structure of the ATPase active form of ORC was not bound to DNA, but it allowed a comparison of the yeast ORC and the human ORC and predicted how the yeast ORC binds in a sequence-specific manner to origin DNA, whereas the human ORC does not. The yeast ORC-Cdc6 structure within the OCCM suggested that there were DNA-specific interactions with an α-helix in Orc4, with a helix and loop in Orc2 and two small loops in Cdc6. Strikingly, these amino acid sequences are absent in the human ORC subunits, and indeed an examination of the conservation of these helices and loops shows that they are restricted to a subset of budding yeasts that have DNA sequence–specific origins of DNA replication. We have mutated the Orc4 α-helix and, using genetic selection methods, we have been able to change the DNA sequence specificity at origins in *S. cerevisiae*, as well as change the distribution of active origins across the yeast genome. A complete analysis of these mutants is currently under way.

Assembly of Proteins on Pre-Replicative Complexes at Origins of DNA Replication

Once the pre-RC has been assembled in the G_1 phase of the cell division cycle, a number of DNA replication initiation activating proteins are assembled onto the Mcm2-7 DH, some of which are dependent on the activity of the cyclin-dependent protein kinase (CDK) and the Dbf4-dependent protein kinase (DDK), two kinases that become active at the G_1 to S phase transition. We have investigated how these proteins interact, focusing on the large platform protein Dpb11. Dpb11 is known to be required for the initiation of DNA replication, but also is involved in monitoring DNA damage at the DNA replication fork as a DNA damage checkpoint protein.

Initially, we expressed these activating proteins using recombinant baculovirus expression in insect cells, including the six ORC subunits, Cdc6, Cdt1, the six Mcm2-7 hexamer subunits, the helicase-associated proteins Cdc45, and the four subunit GINS (Sld5, Psf1, Psf2, Psf3), as well as Dpb11, Sld2, Sld3, and the Sld3 binding protein Sld7, along with the activating kinases CDK and DDK. It is known from research in John Diffley's laboratory that Sld2 and Sld3 bind to the phospho-binding BRCT motifs Dpb11 only when they have been phosphorylated by CDK. We demonstrated that Cd45, GINS, Mcm3, and Cdc45 all bound to the Dpb11 protein independent of CDK or DDK by co-expressing the proteins in insect cells and immunoprecipitation. Those proteins that interacted with Dpb11 were expressed in bacteria in which there is no phosphorylation of the

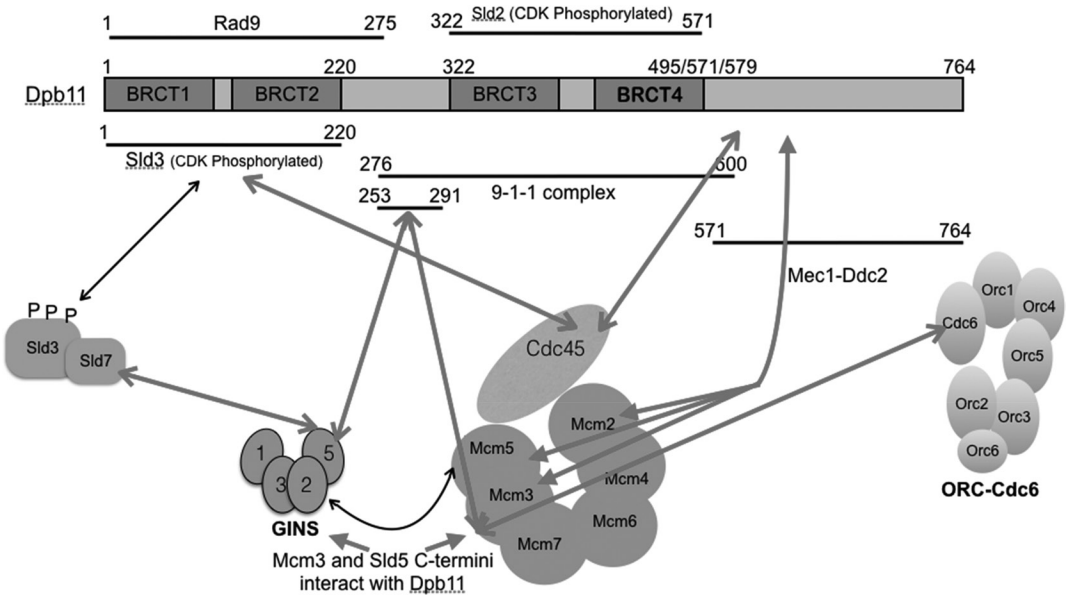


Figure 2. Summary of the protein–protein interactions with Dpb11. Dpb11 functions as a platform for interactions that are required for the initiation of DNA replication. The interactions shown with thicker arrows were identified in this study.

subunits, and the amino acids within Dpb11 and the interacting proteins were mapped. A summary of these interactions is shown in Figure 2.

Cdc45, a component of the CMG (Cdc45, Mcm2-7, GINS) helicase, bound to the carboxyl terminus of Dpb11 and to Sld3. The most interesting of these interactions were the interactions between Mcm3 and GINS with Dpb11. Previously it has been shown that ORC1-6 and Cdc6 bind to DNA to facilitate the recruitment of the Mcm2-7 double hexamer in a process that requires Cdc6 interaction with the carboxyl terminus of Mcm3. We demonstrated that the very same Mcm3 carboxyl terminus that binds to Cdc6 also binds to Dpb11 (yeast ortholog of human TopBP1), the BRCT-containing protein that is involved in the initiation of DNA replication and, separately, checkpoint control of DNA damage. We mapped the Mcm3-C binding site to the same 39 amino acid (aa) region between the BRCT2 and BRCT3 motifs of Dpb11. Interestingly, the carboxyl terminus of the GINS subunit Sld5 binds the same region in Dpb11, and Mcm3 and Sld5 proteins have related carboxy-terminal sequences. Mutations of the carboxyl terminals of Sld5 and Mcm3 do not bind Dpb11. The carboxyl terminus of yeast Mcm3 and Sld5 sequences are more related to each other than the corresponding human Mcm3/Sld5 C-termini, which are themselves similar, suggesting species specificity in interactions with Cdc6 and Dpb11. Indeed, these interactions are not only evolutionarily conserved, but species specific as the 6-aa carboxy-terminal peptide of yeast Mcm3 and Sld5 binds to Dpb11 but fails to bind to TopBP1. Conversely, the 6-aa carboxy-terminal peptide of human MCM3 and GINS4 (Sld5 ortholog in humans) binds to TopBP1 and not to Dpb11, suggesting species specificity at the D/G amino acid of the carboxy-terminal region of these essential replication factors.

The interactions between Sld5 and Dpb11 are regulated by protein phosphorylation of Dpb11, and we found that the structure of the Dpb11 protein is also regulated by protein phosphorylation. Gel filtration, gradient sedimentation analysis, and electron microscopy demonstrated that the phosphorylated and nonphosphorylated Dpb11 proteins are a monomer, but the nonphosphorylated Dpb11 forms an extended linear structure, whereas the phosphorylated Dpb11 forms a circular structure. We are currently searching for the kinase that phosphorylates Dpb11 and changes its structure.

Yeast strains were constructed that express either the wild-type Sld5 protein or a mutant version of Sld5 that has the very carboxy-terminal amino acid changed so that it no longer binds Dpb11. The mutant Sld5 strains grow nearly as well as the wild type in normal media, but they grow poorly in the presence

of the drug hydroxyurea, which inhibits the activity of ribonucleotide reductase and reduces dNTPs in the cell. This suggests that replication is compromised in vivo when replicative stress is applied. In collaboration with Dirk Remus at Memorial Sloan Kettering Cancer Center, we compared the wild-type and mutant Sld5 proteins that had been reconstituted into the GINS complex in vitro, using a reconstituted DNA replication system that can replicate DNA from the yeast origin with highly purified proteins. The results show that the mutant GINS is partially defective in the initiation of DNA replication, but not in the rate of fork progress once replication is under way.

PUBLICATIONS

Noguchi Y, Yuan Z, Bai L, Schneider S, Zhao G, Stillman B, Speck C, Li H. 2017. Cryo-EM structure of Mcm2-7 double hexamer on DNA suggests a lagging strand DNA extrusion model. *Proc Natl Acad Sci* **114:** E9529–E9538.

Tocilj A, Yuan Z, On KF, Sun J, Elkayam E, Li H, Stillman B, Joshua-Tor L. 2017. Structure of the active form of human origin recognition complex and its ATPase motor module. *eLife* 2017;10.7554/eLife.

Yuan Z, Riera A, Bai L, Sun J, Nandi S, Spanos C, Chen ZA, Barbon M, Rappsilber J, Stillman B, Speck C, Li H. 2017. Structural basis of MCM2-7 replicative helicase loading by ORC-Cdc6 and Cdt1. *Nat Struct Mol Biol* **24:** 316–324.

In Press

On KF, Jaremko M, Stillman B, Joshua-Tor L. 2018. A structural view of the initiators for chromosome replication. *Curr Opin Struct Biol* **53:** 131–139.

Stillman B. 2018. Histone modifications: insights into their influence on gene expression. *Cell* **175:** 6–9.

Tramentozzi E, Ferraro P, Hossain M, Stillman B, Bianchi V, Pontarin G. 2018. The dNTP triphosphohydrolase activity of SAMHD1 persists during S-phase when the enzyme is phosphorylated at T592. *Cell Cycle* **17:** 1102–1114.

TRANSCRIPTIONAL VULNERABILITIES OF THE CANCER CELL

C.R. Vakoc Y. Huang, S. Hur, O. Klingbeil, B. Lu, J. Milazzo, S. Polyanskaya, J. Roe, T. Somerville, M. Sroka, Y. Tarumoto, Y. Wei, X. Wu, Y. Xu, Z. Yang

Broad alterations of gene expression are necessary, and often sufficient, for cells to undergo malignant transformation. As a consequence, cancer cells are vulnerable to perturbations of transcriptional regulators, including DNA-binding transcription factors (TFs) and chromatin regulatory machineries. Our laboratory has taken a genetic screening approach to identify essential transcriptional regulators in cancer cells in an effort to expose basic regulatory mechanisms and opportunities for therapeutic intervention. Our initial efforts used short hairpin RNA (shRNA) screening but, more recently, we have embraced the use of CRISPR-Cas9 genome editing as a tool for annotating essential regulators in cancer. The overarching goal of our research is to gain a mechanistic understanding of how transcriptional regulators become essential in specific cancer contexts. Our efforts have historically focused on identifying essential chromatin regulators in the MLL-fusion subtype of acute myeloid leukemia (AML), but we now study four different types of cancer in search of new therapies.

A TFIID Perturbation that Targets Myb and Suppresses Acute Myeloid Leukemia

Y. Xu

Targeting of general coactivators, such as BRD4, is an emerging strategy to interfere with oncogenic TFs in cancer. However, coactivator perturbations have the potential to influence the function of numerous TFs, thereby resulting in biological pleiotropy. We have used an shRNA screen to identify TAF12, an 18-kDa subunit of TFIID and SAGA coactivator complexes, as a selective requirement for AML progression. We have traced this AML-specific dependency to a direct interaction between the TAF12/TAF4 histone-fold heterodimer and the transactivation domain of Myb, a TF with established roles in leukemogenesis. Ectopic expression of a histone-fold domain fragment of TAF4 can efficiently squelch TAF12 in cells, suppress Myb, and regress AML in mice. Our study reveals a strategy for potent Myb inhibition in AML and highlights how an oncogenic TF can be selectively neutralized by targeting a general coactivator complex.

Master Regulators and Enhancer Reprogramming in Pancreatic Ductal Adenocarcinoma

S. Hur, J. Roe, T. Somerville [in collaboration with David Tuveson's laboratory]

Pancreatic ductal adenocarcinoma (PDA) is one of the most lethal human malignancies, owing, in part, to its propensity for metastasis. We have used an organoid culture system to investigate how transcription and the enhancer landscape become altered during discrete stages of disease progression in a PDA mouse model. This approach has revealed that the metastatic transition is accompanied by massive and recurrent alterations in enhancer activity. We implicated the TF FOXA1 as a driver of enhancer activation in this system, a mechanism that renders PDA cells more invasive and less anchorage-dependent for growth in vitro, as well as more metastatic in vivo. In this context, FOXA1 activates a foregut endoderm transcriptional program without altering the expression of genes associated with the epithelial-to-mesenchymal transition. Collectively, this study implicates FOXA1 up-regulation, enhancer reprogramming, and a novel retrograde developmental transition in PDA progression and metastasis.

Another strategy we have taken to understand PDA metastasis is to investigate the transcriptional heterogeneity of PDA tumors between individuals, which can be broadly classified into "progenitor" or "squamous" subtypes. We have initiated studies aimed

at defining the critical master regulators that specify these two disease subtypes, with a focus on squamous cell identity. We are learning that squamous identity is covered by two TFs, which cooperate to promote the most aggressive form of this disease. In the longer term, these efforts are aimed at exposing therapeutic targets in these two subtypes of disease and with relevance to metastatic disease.

A New Cell-of-Origin and Molecular Subtype of Small-Cell Lung Cancer

Y. Huang, X. Wu

Small-cell lung cancer (SCLC) is one of the deadliest human malignancies, with a paucity of oncogenic targets for therapeutic intervention. We have applied domain-focused CRISPR screening to identify TF dependencies in SCLC with exceptional potency and specificity for this tumor context. This approach led us to the lineage-specific TF POU2F3, which is expressed and essential in a subset of human SCLC lines. Using patient sample-derived transcriptome data, we have shown that POU2F3 expression is exclusive to a variant subtype of SCLC characterized by low expression of neuroendocrine differentiation markers. Functional and epigenomic experiments implicate POU2F3 as a master regulator of this tumor cell state. CRISPR screening also revealed that POU2F3$^+$ SCLC lines are hypersensitive to targeting of the insulin-like growth factor 1 receptor (IGF1R) and targeting of SOX9 and ASCL2. Together, these findings suggest that POU2F3$^+$ SCLC is a diagnostic entity associated with unique transcriptional and signaling vulnerabilities.

Targeting Fusion Oncoproteins in Leukemia and Sarcoma

B. Lu, S. Polyanskaya, M. Sroka, Y. Tarumoto, Z. Yang

Chromosomal translocations that fuse the amino terminus of MLL with various carboxy-terminal partner proteins are powerful genetic drivers of acute leukemia. We have used domain-focused CRISPR screening to identify zinc finger protein 64 (ZFP64) as an essential TF in human leukemia lines harboring *MLL* translocations. Using RNA-seq, we show that ZFP64 functions in a common pathway with endogenously expressed MLL fusion proteins. However, we found, unexpectedly, that ZFP64 is dispensable in engineered leukemia lines that express an MLL fusion cDNA from a retroviral promoter. To explain this observation, we characterized the sequence-specific DNA-binding function of ZFP64. We found that the *MLL* proximal promoter is an outlier in the human genome with regard to its density of ZFP64 motifs, levels of ZFP64 occupancy, and dependence on ZFP64 for expression. Our results suggest that ZFP64 is required to express the MLL fusion from its endogenous chromosomal context, thus demonstrating how a transcriptional dependency can be imposed by an oncogene's promoter sequence.

The lineage-specific TF MEF2C is commonly deregulated in leukemia. However, strategies to target this TF have yet to be identified. We have used a domain-focused CRISPR screen to reveal an essential role for LKB1 and its salt-inducible kinase effectors (SIK3, in a partially redundant manner with SIK2) to maintain MEF2C function in AML. A key phosphorylation substrate of SIK3 in this context is histone deacetylase 4, a repressive cofactor of MEF2C. Consequently, targeting of LKB1 or SIK3 diminishes histone acetylation at MEF2C-bound enhancers and deprives leukemia cells of the transcriptional output of this essential TF. We also found that MEF2C-dependent leukemias are sensitive to on-target chemical inhibition of SIK activity. This study reveals a chemical strategy to block MEF2C function in AML, representing an opportunity for biomarker-based drug discovery in this disease.

CRISPR screening efforts in the lab continue to uncover remarkable transcriptional and signaling vulnerabilities in MLL-rearranged leukemia. In addition to the targets above, we have also begun investigating a nuclear serine phosphatase, as well as ubiquitin E2/E3 pair, as being unique vulnerabilities in this disease. Ongoing mechanistic studies will reveal the underlying role of these factors in this context.

Alveolar rhabdomyosarcoma (aRMS) is a rare muscle cancer that affects primarily children and adolescents. Although the disease bears a low overall mutational burden, >60% of aRMS patients harbor the t(2;13)(q35;q14) translocation that leads to expression of the PAX3-FOXO1 fusion oncoprotein, in which the DNA-binding domain of PAX3 is linked to the transactivation domain of FOXO1. Fusion-positive tumors are the most aggressive, with a 4-year overall survival

rate of 8% for metastatic tumors. Numerous studies showed that aRMS cancers are dependent on the presence of the PAX3-FOXO1 chimera, and withdrawal causes cell death or differentiation. However, no therapies targeting the fusion protein exist to date, partly as a result of challenges associated with designing drugs that target TFs. The goal of our ongoing efforts is to elucidate the entire PAX3-FOXO1 fusion oncoprotein pathway in aRMS. To this end, we are characterizing how aRMS cells respond to fusion inactivation, as well as the molecular signatures and dynamics of each response. We are developing an unbiased, reporter-based, FACS-assisted CRISPR screening strategy to identify genes that phenocopy fusion inactivation on knockout. These efforts will impact our understanding of the basic biology of fusion-positive aRMS. Further, the findings might allow the development of therapeutics that indirectly silence the pathway by interfering with factors known to cooperate with the fusion oncoprotein in the maintenance of the transformed state.

Next-Generation CRISPR Screening Strategies

O. Klingbeil, Y. Wei [in collaboration with O. El Demerdash]

One limitation of our current CRISPR screening strategy is that we are only inactivating a single gene in our screens. Because many genes possess highly similar homologs, our approach cannot expose redundant gene relationships. To address this, we are generating a scalable method for cotargeting of homologous domains using CRISPR-CPF1. In a second area of technology development, we seek to probe essential transcriptional regulators and signaling molecules in the in vivo context. For this purpose, we have constructed a robust genome editing approach to defining in vivo requirements for AML progression.

PUBLICATIONS

Barbieri I, Tzelepis K, Pandolfini L, Shi J, Millán-Zambrano G, Robson SC, Aspris D, Migliori V, Bannister AJ, Han N, et al. 2017. Promoter-bound METTL3 maintains myeloid leukaemia by m^{6}A-dependent translation control. *Nature* **552:** 126–131.

Fontanals-Cirera B, Hasson D, Vardabasso C, Di Micco R, Agrawal P, Chowdhury A, Gantz M, de Pablos-Aragoneses A, Morgenstern A, Wu P, et al. 2017. Harnessing BET inhibitor sensitivity reveals AMIGO2 as a melanoma survival gene. *Mol Cell* **68:** 731–744.

Ipsaro JJ, Shen C, Arai E, Xu Y, Kinney JB, Joshua-Tor L, Vakoc CR, Shi J. 2017. Rapid generation of drug-resistance alleles at endogenous loci using CRISPR-Cas9 indel mutagenesis. *PLoS ONE* **12:** e0172177.

Lanning BR, Vakoc CR. 2017. Single-minded CRISPR screening. *Nat Biotechnol* **35:** 339–340.

Roe JS, Hwang CI, Somerville TDD, Milazzo JP, Lee EJ, Da Silva B, Maiorino L, Tiriac H, Young CM, Miyabayashi K, et al. 2017. Enhancer reprogramming promotes pancreatic cancer metastasis. *Cell* **170:** 875–888.

Vakoc CR, Tuveson DA. 2017. Soils and seeds that initiate pancreatic cancer metastasis. *Cancer Discov* **7:** 1067–1068.

Vakoc CR, Tuveson DA. 2017. Untangling the genetics from the epigenetics in pancreatic cancer metastasis. *Nat Genet* **49:** 323–324.

Xu Y, Vakoc CR. 2017. Targeting cancer cells with BET bromodomain inhibitors. *Cold Spring Harb Perspect Med* **7:** a026674.

CANCER: GENETICS

Christopher Hammell's lab is interested in understanding gene regulatory processes that give rise to robust phenotypes associated with normal development in animals (specifically, how the timing of developmental processes is controlled), as well as the alterations in these pathways that give rise to diseases such as cancer (as in the alterations in mitogenic pathways in melanoma). Hammell and colleagues approach this elemental problem by using a variety of model organisms and patient-derived cancer cell lines. To directly identify the components that function in controlling normal developmental timing, they use the small nematode *Caenorhabditis elegans*, applying forward and reverse genetic approaches. In contrast to the extreme robustness of cell-fate lineage in *C. elegans*, in which specification of developmental programs is hardwired, mutations that alter conserved signaling pathways in melanoma create relatively plastic developmental landscapes that allow these lesions to become aggressive tumors. Notably, the gene regulatory architecture of melanoma cells allows them to acquire resistance to therapeutic agents. Hammell's team is interested in epigenetic mechanisms that contribute to resistance, specifically dramatic changes in gene expression patterns and intracellular signaling pathways. They are performing high-throughput screens to identify cellular factors that allow these rewiring events to occur, with the idea that these components would make ideal therapeutic targets to complement existing clinical strategies.

The **Je H. Lee** laboratory studies how cells interact with their microenvironment to regulate gene expression during development. Single-cell heterogeneity in gene expression can result from spatial differences in cell–cell and cell–extracellular matrix (ECM) interactions. Such differences contribute to stochastic evolution of tumor cells, as well as morphogenesis during normal development. However, the spatial control of gene expression in complex tissues, embryos, or tumors remains poorly understood because most genome-wide studies sample bulk tissues or dissociated single cells.

The Lee lab has recently developed a method to sequence RNA molecules directly within single cells and tissues using subcellular resolution imaging, and they showed subtle differences in cell–cell/ECM signaling and gene expression genome-wide in situ. By clustering transcripts into functionally or morphologically discrete regions, we find many unique spatial markers and signaling pathways. The lab focuses on the role of noncoding RNA in chromatin remodeling and tumor progression using mouse and organoid models of human cancer. The team uses in situ sequencing, cell lineage tracing, and single-cell profiling to understand how noncoding RNA affects the evolution of tumor cells in their native context. The lab's long-term goal is to develop better tumor classification tools and anticancer therapeutics using our understanding of the tumor microenvironment.

Alea Mills is studying genetic pathways important in cancer, aging, and autism, identifying the genetic players and determining how aberrations in their functions culminate in human disease. Through innovative use of a technique called "chromosome engineering," the Mills group discovered that one of the most common genetic alterations in autism—deletion of a 27-gene cluster on chromosome 16—causes autism-like features in mice. These autism-like movement impairments can be identified just days after birth, suggesting that these features could be used to diagnose autism. Mills has also used chromosome engineering to identify a tumor suppressor gene that had eluded investigators for three decades. The gene, called *Chd5*, was shown by Mills to regulate an extensive cancer-preventing network. This year, the Mills lab uncovered how *Chd5* acts as a tumor suppressor: It binds to a protein found within chromatin to turn specific genes on or off, halting cancer progression. The epigenetic role of *Chd5* in development, cancer, and stem-cell maintenance is currently being investigated. The Mills lab is also studying p63 proteins, which regulate development, tumorigenesis, cellular senescence, and aging in vivo. They succeeded in

halting the growth of malignant tumors by turning on production of one of the proteins encoded by the *p63* gene, called TAp63. TAp63 also exerts other protective effects. This year, the Mills lab generated a mouse model that allowed them to find that TAp63 is required to prevent a genetic disorder, known as EEC (ectrodactyly-ectodermal dysplasia cleft lip/palate syndrome), which is characterized by a cleft palate and major deformities of the skin and limbs in infants. In addition, they recently discovered that a different version of *p63*, called ΔNp63, reprograms stem cells of the skin to cause carcinoma development—the most prevalent form of human cancer. Modulation of these proteins may offer new ways to treat human malignancies in the future.

Human cancers show a diverse array of genomic gains and losses that alter the dosage of hundreds of genes at once. About 90% of solid tumors display whole-chromosome aneuploidy, whereas many tumors with diploid karyotypes nonetheless harbor segmental or arm-length aneuploidies that also result in significant gene copy number alterations. Despite the prevalence of aneuploidy in cancer, its functional consequences for cell physiology remain poorly understood. **Jason Sheltzer** and colleagues have shown the existence of several surprising phenotypes that are shared among cells with different chromosomal imbalances. They have demonstrated that aneuploidy can function as a novel source of genomic instability, as aneuploid cells tend to display elevated levels of mutation, mitotic recombination, and chromosome loss. They have also identified a transcriptional signature of aneuploidy that is associated with cellular stress and slow proliferation and is found in aneuploid primary and cancer cells across a host of organisms. More recently, they have investigated the link between aneuploidy and cellular transformation. Using a series of genetically matched euploid and aneuploid cell lines, they have shown that aneuploidy can paradoxically function as a barrier to tumor growth. They are currently continuing their investigation of the role in aneuploidy in cancer. They are also applying CRISPR-Cas9-mediated genome engineering to develop novel mouse models to explore the impact of gene dosage alterations on tumor development in vivo.

Although aneuploidy is a ubiquitous feature of human tumors, it occurs rarely in somatic cells. Thus, differences between aneuploid and euploid cells may represent crucial therapeutic vulnerabilities in cancer. By identifying phenotypes that are shared among tumors with different aneuploidies, the members of the Sheltzer lab hope to discover pathways that can be manipulated to selectively eliminate aneuploid cells or to block aneuploidy's non–cell autonomous effects. Drugs that target these pathways may have broad utility against a wide range of aneuploid cancers while showing minimal toxicity in euploid tissue.

Michael Wigler's work provides a new paradigm for understanding and exploring human disease. The Wigler lab studies human cancer and the contribution of new mutation to genetic disorders. The cancer effort (with James Hicks, Alex Krasnitz, and Lloyd Trotman) focuses on breast and prostate cancers. It involves collaborative clinical studies to discover mutational patterns predicting treatment response and outcome and the development of diagnostics to detect cancer cells in bodily fluids such as blood and urine. The major tools are single-cell DNA and RNA analysis. The single-cell methods, which are in development, are also being applied to problems in neurobiology (with Josh Huang and Pavel Osten) to characterize neuronal subtypes, somatic mutation, and monoallelic expression. The Wigler lab's genetic efforts are a collaboration with Ivan Iossifov and Dan Levy, and this team focuses on determining the role of new mutations in pediatric disorders. In a large-scale population sequencing project with W. Richard McCombie and the Genome Sequencing Center at Washington University in St. Louis, and supported by the Simons Foundation, the team has proven the contribution of this mechanism to autism. The work further suggests a relationship between the mutational targets in autism and the process of neuroplasticity that lies at the heart of learning. Smaller-scale population studies of congenital heart disease and pediatric cancer (in collaboration with scientists at Columbia University and Memorial Sloan Kettering Cancer Center, respectively) also point to new mutation as a causal factor in these disorders.

GENETIC DISSECTION OF GENES THAT MODULATE TEMPORAL GENE REGULATION

C.M. Hammell C. Aguirre-Chen J. Buonfiglio K. Hills-Muckey J. Wang R. Weinmann

The long-standing question that the lab has focused on in the past few years is how the precise timing of developmental gene regulation is established, in the most basic sense, and how these patterns of gene expression are maintained in highly variable environments. These efforts originated through the genetic dissection of the gene regulatory pathways that establish temporal cell-fate specification in *Caenorhabditis elegans*, and have historically been centered on how the activities of specific microRNAs (miRNAs) impact these processes. Although initially centered on integrating how miRNA-mediated gene regulation is integrated in temporal transitions, our current work has expanded to two major themes: (1) understanding how oscillatory gene expression (both protein coding and regulatory RNAs) is established during postembryonic development, and (2) how a novel and conserved family of proteins modulate translational output during development.

Identification of New Components of the Developmental Oscillator

J. Buonfiglio, K. Hills-Muckey, J. Wang

Previous work from our laboratory has determined that a large fraction of the *C. elegans* transcriptome is expressed in an oscillatory fashion. Animals show periodic patterns of gene expression throughout larval development, and these patterns are tied to the essential molting cycle. Importantly, misregulation of this patterned expression leads to cell fate specification defects in which normal temporal cell fate specification events are precociously executed or inappropriately reiterated. Through the identification of *lin-42*, the *C. elegans* Period ortholog, as a component that controlled oscillatory gene expression, we gained a foothold into the gene regulatory network (GRN) that controls these transcriptional programs. Specifically, mutations in *lin-42* result in an overaccumulation of transcripts whose transcription is paired with the molting cycles, causing these patterns of gene expression to show elevated amplitudes. We, therefore, interpreted the function of *lin-42* in this GRN as a negative regulator of cyclical gene expression. As a consequence of this molecular phenotype, *lin-42* mutant animals show later developmental expression programs earlier than wild-type (i.e., they are precocious). We then searched for genes that also function in this pathway by suppressing the precocious phenotypes shown in animals that harbor *lin-42* mutations. One of the major discoveries our lab has made using this approach was the identification of a number of transcription factors that genetically antagonize normal LIN-42 function.

In the past year, we have focused on the function of one of these specific transcription factors, NHR-23, for a variety of reasons. First, genome-wide chromatin immunoprecipitation–sequencing (ChIP-seq) data from staged animals indicate that NHR-23 binding sites are enriched in the promoters of genes that show oscillatory expression. Putative regulatory targets of NHR-23 include many heterochronic miRNA genes, as well as genomic regions encoding both NHR-23 and LIN-42 (Fig. 1). Second, *nhr-23* mRNA expression is also cyclical, suggesting that its activity as a transcription factor may implicitly generate periodic transcriptional pulses of target genes. Analysis of *nhr-23* mutations and RNA interference (RNAi) phenotypes indicates that it functions to control the repetitive molting cycles of larval development, and genetically interacts with components of the heterochronic pathway. Consistent with a direct role of NHR-23 in promoting normal temporal patterning, RNAi depletion of *nhr-23* phenocopies heterochronic timing defects associated with hypomorphic alleles of the *let-7* family of miRNAs, and is required to limit the supernumerary molting cycles of retarded heterochronic

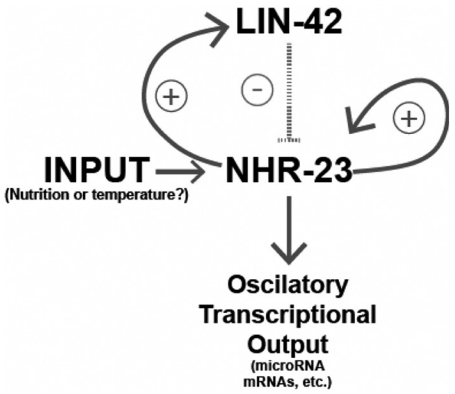


Figure 1. A genetic model for the regulatory architecture that couples gene expression to the rate of animal growth. Green lines indicate transcriptional activities, whereas red lines indicate interactions mediated at the protein level.

mutants (Fig. 2). Finally, RNAi inactivation of *nhr-23* expression has been shown to limit the ability of nutrient-rich food sources to accelerate development, suggesting it may be a component that controls both transcription and normal developmental pace in a ligand-dependent manner.

The hypothesis that NHR-23 and LIN-42 function together to both mediate temporal aspects of developmental transcription *and* the overall pace of animal growth leads to an attractive hypothesis in which these two transcription factors form the central components of a developmental oscillator. In other systems, PERIOD (the human, fly, and mouse ortholog of LIN-42) negatively regulates the activity of two transcription factors, Clock (Clk1) and Bmal (Bmal1), that drive periodic transcription. In this capacity, Period proteins (Per1/Per2) physically associate with Clk1 and Bmal1 to attenuate transcriptional activation. An important feature of this interaction is that, by inhibiting the activity of Clk1 and Bmal1 proteins, Period negatively regulates its own oscillatory transcription, as well as the transcription of Bmal1 and Clk1 genes. This regulatory feature, composing a positive and negative autoregulatory feedback loop, establishes rhythmic circadian expression patterns. We hypothesize that NHR-23 and LIN-42 are integrated into a similar regulatory framework (Fig. 1a). In contrast to coupling transcriptional patterns to daily light and dark cycles, we hypothesize that this unique regulatory framework integrates environmental information to control developmental pace.

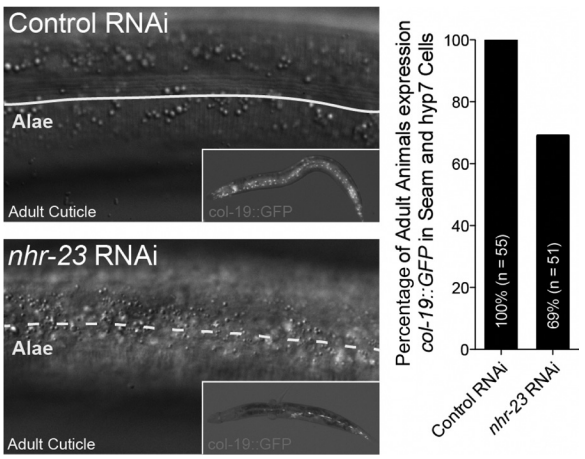


Figure 2. *nhr-23* RNA interference (RNAi) results in heterochronic phenotypes. *nhr-23* RNAi prevents animals from expressing adult-specific alae structures (large image) and expressing adult-specific green fluorescent protein (GFP) reporters (*inset*).

We have tested two aspects of this model. In this model, NHR-23 expression would generate its own expression in an autoregulatory fashion dependent on NHR-23-binding sites found in its promoter. As NHR-23 levels increase, NHR-23 would additionally promote the expression of *lin-42*. This increased expression of LIN-42 would lead to the dampening of expression of both *nhr-23* and *lin-42* genes (Fig. 1). To characterize the temporal expression patterns of NHR-23 and LIN-42 in detail, we quantified changes in the expression of fluorescently tagged translational fusions of each transcription factor in the hypodermal tissues of L4-staged larvae (Fig. 3). These results indicate that the pulse of NHR-23 expression occurs early in each larval period and precedes the onset of LIN-42 expression (Fig. 3A). The temporal arrangement of these expression patterns (order of expression and overlap) are consistent with our previously described roles for LIN-42 in dampening the amplitude of genes that show pulsatile expression patterns. Second, the model outlined in Figure 1A posits that LIN-42 modulates the transcriptional activities of NHR-23. We, therefore, sought to determine if NHR-23 and LIN-42 physically associate in tissues in which we can describe genetic/regulatory roles for these proteins. For these experiments, we used a bimolecular fluorescence complementation (BiFC)/"split green fluorescent protein" (GFP) system to determine *when* and *where* these proteins may interact in developing larvae. In these assays, translational fusions of candidate

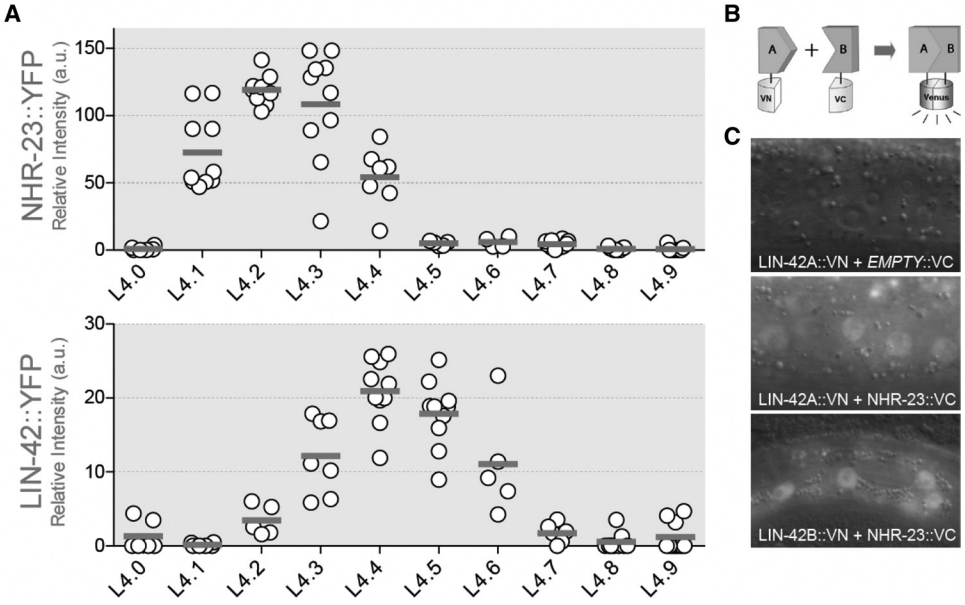


Figure 3. NHR-23 and LIN-42 are expressed at different phases of each larval stage. (*A*) Expression of the nuclear hormone receptor NHR-23::YFP (yellow fluorescent protein) precedes the expression of the negative regulatory component, LIN-42, during the fourth larval stage. (*B*) A schematic outlining how the physical interactions between candidate proteins, A and B, facilitate the complementation of a split Venus fluorescent reporter. (*C*) The results of expression of various bait and prey combinations that show that NHR-23 and LIN-42 proteins are in close proximity in the hypodermal tissues of developing animals.

interacting proteins are generated with each partner expressing a nonfluorescent fragment of GFP. If two proteins are in proximity in vivo, close physical association between the partners allows the tethered GFP fragments to generate a functional, fluorescent GFP structure (Fig. 3B). As shown in Figure 3C, our experiments indicate that both LIN-42 isoforms interact with NHR-23 in hypodermal nuclei. These data suggest that a common gene regulatory architecture similar to the one that generates oscillatory expression during circadian gene regulation may function during *C. elegans* larval development. We hypothesize that features of this developmental clock are modulated by the environment, leading to changes in periodicity, amplitude, phasing, etc. in the transcription of downstream targets that mediate aspects of development.

We are currently developing a series of auxin-induced-degradation alleles of NHR-23 that will allow us to dissect the molecular function of NHR-23 in directing oscillatory transcription. Furthermore, we are using heterologous expression systems to determine if NHR-23 and LIN-42 directly interact. Finally, we are using a variety of GFP-pest reporters for the *nhr-23* and *lin-42* genomic regulatory regions to determine if *nhr-23* promotes its own expression, as well as the expression of its hypothetical negative regulator, *lin-42*, as outlined in Figure 1.

PQN-59, and Its Human Orthologs UBAP2 and UBAP2L, Regulate Translational Output

C.M. Hammell, R. Weinmann

One of the longest-standing projects in our laboratory is the characterization of *pqn-59*, encoding a putative prion domain-containing protein, which we also identified as a suppression of various miRNA mutants. In the past, we have thoroughly characterized the genetics of *pqn-59* both as a suppressor of heterochronic mutants and as a gene that is essential for normal germ line function. We have now focused on discerning the biochemical elements of PQN-59 function in vitro and also moved to study the function of two human PQN-59 orthologs, UBAP2 and UBAP2L.

Analysis of the primary sequence of the PQN-59 protein indicates that it harbors a statistical

overrepresentation of asparagine and glutamine residues in the carboxy-terminal domain (Fig. 4A). In addition to harboring rare amino acids, PQN-59 is a very abundant protein as measured by whole-animal proteomics (Fig. 4B). Other proteins containing these features (i.e., LC domains) are predicted to show unique biochemical/physical properties. Among these is an ability to self-assemble in liquid solution and phase separate. To determine if PQN-59 harbors LC domains, we used a chemical reagent that specifically precipitates LD domains (B-isox). B-isox precipitates the endogenous protein from whole worm lysates, as

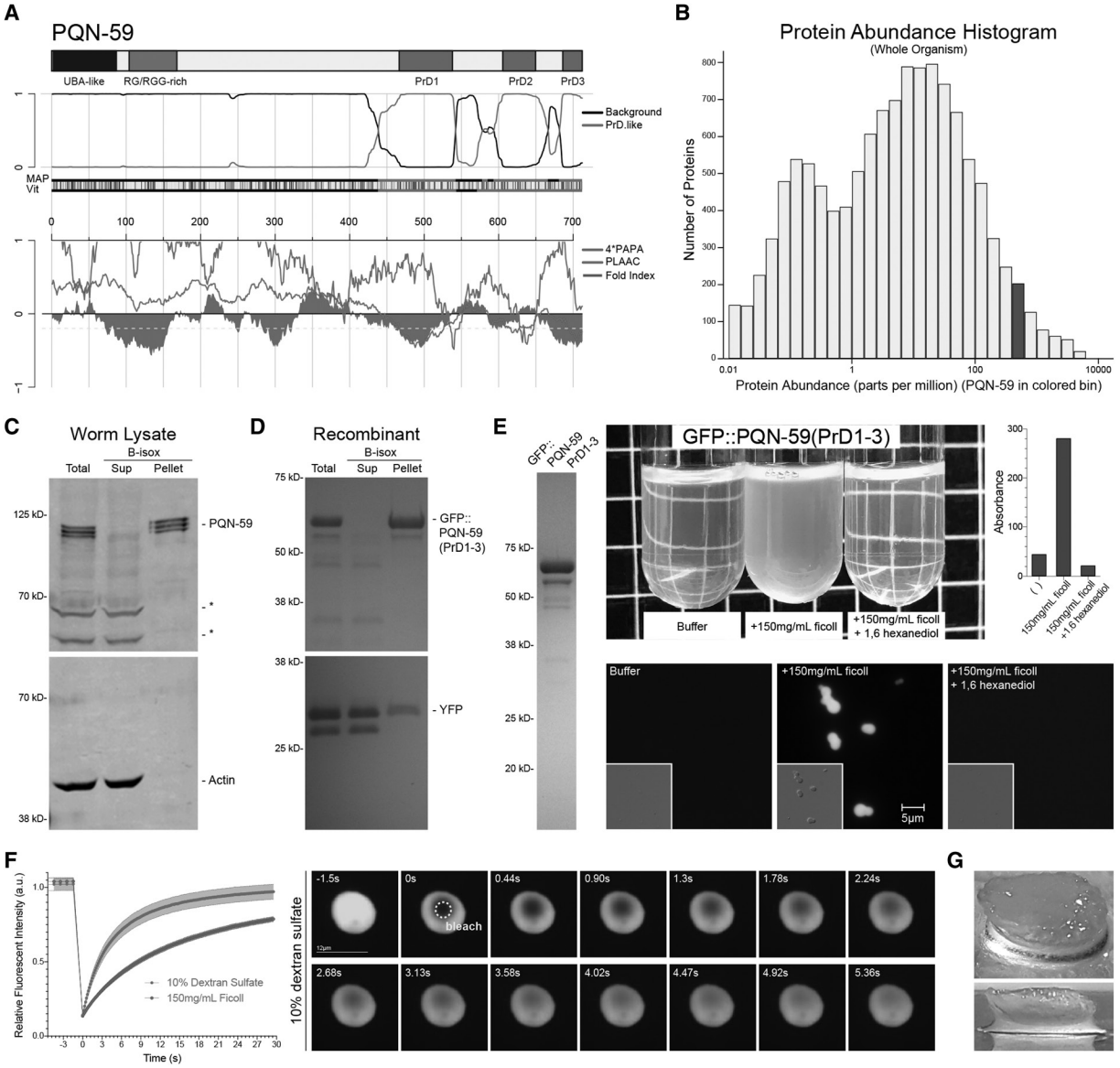


Figure 4. PQN-59 encodes an abundant *Caenorhabditis elegans* protein that shows unique biochemical properties. (*A*) A schematic diagram of the PQN-59 protein domain structure. (*B*) Whole animal proteomics indicate that PQN-59 is a very abundant protein. (*C,D*). The B-isox compound that preferentially precipitates low-complexity proteins precipitates PQN-59 from whole worm extracts and recombinant PQN-59(PrD1-3). (*E*) Recombinant PQN-59(PrD1-3) shows liquid–liquid phase separation in vitro. (*F*) The second phase of PQN-59(PrD1-3) proteins show a liquid state as measured by fluorescence recovery after photobleaching (FRAP). (*G*) Extended incubation of PQN-59(PrD1-3) leads to a gel-like form of the protein.

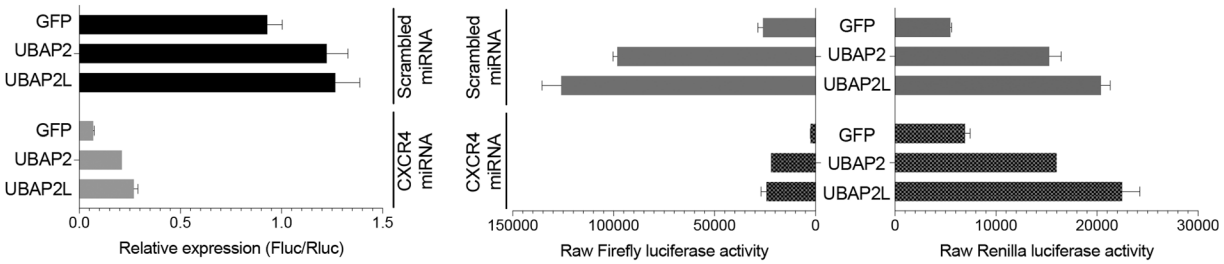


Figure 5. Overexpression of the human orthologs of PQN-59, UBAP2 or UBAP2L, stimulates the translation of reporter transgenes.

well as a recombinant purified fusion protein containing the predicted LC/"prion-like" carboxy-terminal domains (Fig. 4C,D). We next determined if the LC domains of PQN-59 showed liquid–liquid phase separation by concentrating the recombinant protein with a variety of condensing reagents. These experiments show that the prion domains of PQN-59 can phase separate in vitro (Fig. 4E). Importantly, the liquid droplets of PQN-59 are truly liquid matter, as fluorescence recovery after photobleaching (FRAP) indicates that they flow in vitro.

Our genetic analysis indicated that PQN-59 functions to antagonize miRNA-mediated repression. This activity could occur through a variety of molecular mechanisms. To dissect these in more detail, we determined how the human versions of these proteins function. In simple reporter experiments in which we could control the miRNA-mediated repression of a target transcript experimentally, we could show that the human orthologs, UBAP2 and UBAP2L, functioned in a similar manner (i.e., overexpression of UBAP2 or UBAP2L led to a derepression of the miRNA target) (Fig. 5). Subsequent analysis of these activities indicates that overexpression of UBAP2 or UBAP2L increases translation in general (Fig. 5). Future experiments will be aimed at determining how UBAP2 and UBAP2L modulate translational output.

SPATIAL ORGANIZATION OF GENE AND EPIGENETIC REGULATION

J. Lee D. Furth S. Weinmann
D. Ghosh X. Yuan
E. Rozhkova

Our long-term goal is to create tools capable of spatially mapping transcriptional, epigenetic, and posttranscriptional events genome-wide, including often-rare RNA transcripts, nucleotide variants, and nucleic acid modifications, for three-dimensional (3D) reconstruction in human development and cancer. We hypothesize that the functions of some human noncoding RNAs and RNA modifications are specific to higher primates and regulate recently evolved organismal traits. To achieve this aim, we believe that unbiased *spatial* mapping of RNA variants or modification is necessary to directly link gene regulation to form and function to advance understanding of their molecular function or evolutionary role in human tissues.

Nucleolar RNA and Epigenetics in Cancer

E. Rozhkova, D. Ghosh

We are investigating a link between cellular quiescence, cell plasticity, and epigenetics in cancer. Isocitrate dehydrogenase (*IDH*) mutations are found in up to 70% of low-grade glioma (LGG) in patients. Mutations to α-thalassemia/mental retardation syndrome X-linked (*ATRX*) and deletions to *1p19q* occur in a mutually exclusive manner, leading to more aggressive astrocytoma or treatment-sensitive oligodendroglioma, respectively. Both types of tumors are slow-growing and infiltrative and often associated with increased stem cell markers and decreased cell differentiation. Although *IDH* mutations have been shown to cause genome-wide hypermethylation and epigenetic abnormalities, a mechanistic explanation of how *IDH*, *ATRX*, and *1p19q* genetic alterations cooperate to promote tumor progression, alter treatment responsiveness, or produce LGG tumor behavior has been lacking.

We hypothesize that *IDH* mutation leads to increased cell growth after overcoming the initial suppressive effect induced by *IDH* mutation or 2-hydroxyglurate (2-HG). Because *ATRX/p53* and *1p19q/TERT* alterations are mutually exclusive, we hypothesize that they belong to a genetic pathway affected by *IDH* mutation. Hypermethylation induced by *IDH* mutation has been shown to alter CTCF binding to the chromatin, and *ATRX* mutation is linked to hypomethylation of ribosomal DNA (rDNA) loci in the nucleolus. In fact, *1p19q* contains a known regulator of an important nucleolar protein, nucleophosmin (NPM), and p53 and TERT regulate Pol I–dependent transcription in the nucleolus. Given these relationships, we asked whether the rDNA loci (200–400 repeats across five acrocentric chromosomes) have altered CTCF binding in response to *IDH*, *ATRX*, or *1p19q* alterations.

We find that CTCF binding to rDNA is enriched in the Pol I–transcribed region (47S rRNA) rather than the insulator region between repetitive rDNA elements. In addition, this pattern was specific to primary human astrocytes when compared with other human tissue types. Moreover, "immortalization" of human astrocytes using viral oncogenes led to the complete loss of CTCF binding to rDNA. Interestingly, ChIP-seq experiments indicate that CTCF binding overlaps Pol II binding to the 47S rRNA gene body, especially in the 5′ ETS region known to generate antisense RNAs. Because small noncoding RNAs (including antisense RNAs) transcribed from rDNA regulate the epigenetic state in quiescent cells (Martienssen Lab, CSHL), we hypothesize that the role of *ATRX* and *1p19q* genes might be related to the heterochromatin regulation in response to astrocyte quiescence. In fact, recent studies have shown that *IDH* mutation or *2-HG* has a direct effect on cellular energy production, independent of its effect on DNA methylation. Our preliminary data indicate that 2-HG can induce autophage in primary astrocytes, but not in immortalized astrocytes, providing a functional assay to investigate genetic and molecular interactions among *IDH*, *ATRX*, *1p19q* genes (*CIC*, *FUBP*), and *NPM*.

Our preliminary model is as follows: The initial IDH mutation and subsequent 2-HG accumulation create an intracellular environment similar

to a low energy state, activating an adult stem-cell-like quiescence program involving rDNA heterochromatinization, epigenetic remodeling, and autophagy-associated homeostasis. We believe that *ATRX* or *FUBP* mutation decouples the growth inhibitory effect of 2-HG while maintaining the chromatin or cellular plasticity, allowing for infiltrative tumor cell invasion and progression. Unfortunately, the entry or exit from quiescence cannot be recapitulated in vitro for human primary astrocytes. In addition, molecular markers and methods for identifying or isolating quiescent astrocytes from human tissues do not exist. We now have a way to sequence single DNA or RNA molecules directly inside intact cells for amplification, imaging, or cell sorting (termed seqPCR). Our goal for the coming year is to apply seqPCR to isolate human astrocytes using rDNA-derived small RNAs to investigate the correlation between rDNA methylation and gene expression. Fundamentally, this will establish a new experimental paradigm for studying rare cell types or states that lack traditional biomarkers (e.g., cell surface antigens), especially in human tissues.

Self-Organization of Cell Oscillation and Fate Specification

X. Yuan, D. Furth, D. Ghosh

Classic models of biological patterning rely on diffusible molecular or chemical cues between cells; however, competing mechanical forces or oscillating cell states within a population can also generate similar patterns. In fact, "waves" of gene expression can be observed in many developing systems, and they represent a collective action of individual cells that oscillate in phase or out of phase with one another. We hypothesize that short- or long-range cell–cell interactions associated with the density-dependent activation or inhibition of cell growth can generate spatially organized "waves" during embryonic tissue patterning. Recently, studies have shown that mid-G_1 regulators (e.g., Rb, cdk4/6, cyclin D) can control gastrulation-associated gene expression programs in human embryonic stem cells (hESCs). Furthermore, other groups have shown that the cell density, population size, and regional geometry are critical determinants of primary germ layer patterning in vitro. Although mid-G_1 transition defects are associated with abnormal S-phase entry in cancer, this view does not explain the emergence of multiple pocket protein members (Rb, p130, p107), cyclins D1-3, E2F1-6, cdk4/6, and other chromatin remodelers during the evolution of higher mammalian organisms.

We hypothesize that direct cell–cell interaction is stimulatory for cell growth, whereas long-range mechanotransduction mediates lateral inhibition of cell proliferation. Based on these properties in hESCs, we expect that the rate of G_1 progression can vary in space, leading to self-organization or patterning of cell states capable of responding to specific growth factors, if suitable topological or boundary parameters are present. Using hESCs labeled with fluorescent cell cycle indicators for live cell imaging, other groups and we have observed "spots," "rings," or "waves" in differentiation-primed hESCs, if the population size and density reach a certain threshold. We hypothesize that noncanonical mid-G_1 regulators are in fact oscillating regulators of gene expression associated with transcriptional competency during cell fate specification. Our approach is to focus on neuroectodermal cell fate specification of hESCs during gastrulation-like gene expression reprogramming in mid-G_1 in vitro. Our first aim is to map mid-G_1 and associated chromatin regulators genome-wide during specific phases of the cell cycle in undifferentiated hESCs. Our next aim is to focus on neuroectodermal gene expression and identify cell fate–switching mid-G_1 complexes responsive to cell cycle oscillations. Finally, we plan to investigate how mechanotransduction signaling (e.g., Rho, YAP) or sensors (e.g., cadherins, actin) regulate such complexes during the spatial patterning of cell fate.

To observe whether cell state or transcriptional competency patterns are present in vivo, we are developing highly sensitive in situ RNA-sequencing technologies to reconstruct the whole transcriptome in self-organizing hESCs. Our first goal is to benchmark the sensitivity and specificity of our technology using monoallelic gene expression patterns in newly established X-chromosomal inactivation in hESCs. We will sequence the transcriptome of multiple self-organizing hESC clusters in parallel to refine and reconstruct their gene expression patterns in 3D, and we will also integrate single-cell RNA-sequencing (RNA-Seq) data to map specific cell subpopulations and their microenvironmental niches within the emerging tissue in situ.

Technology Development for Mapping Gene Regulation in Tissues

D. Furth, D. Ghosh, S. Weinmann, X. Yuan, E. Rozhkova

Image-based RNA quantification methods are commonly used for validating a select number of gene expression markers in the tissue because most existing methods are not scalable for comprehensive profiling of genetic or epigenetic signatures de novo. In addition, they lack the single-nucleotide specificity and sensitivity for extensively quantifying mutations, polymorphisms, epigenetic modifications, posttranscriptional alterations, or small RNAs. Regardless of technical specifications, however, transformative technologies must address the following: (1) adoptability across a broad range of applications, (2) unbiased or comprehensive molecular profiling, (3) accessibility and simplicity of implementation, and (4) standards for measuring performance in biologically relevant systems.

We are developing new RNA template-based in situ sequencing chemistry that enables compact oligonucleotide designs able to directly sequence mono- or polynucleotide RNA variants. The sequencing reaction product tags endogenous RNA variants, initiating isothermal polymerase-based linear amplification for sequential signal readout and image reconstruction. Applications include (1) iteratively mapping a large number of genes or expression modules in 3D, (2) visualizing allele-specific expression (ASE) in complex traits, (3) cataloging tumor mutations to track cancer cell evolution, (4) sequencing genetic alterations to reconstruct cell lineages, and (5) isolating cells based on DNA or RNA variants for genome-wide sequencing studies.

Our short-term goal is to focus specifically on benchmarking our in situ RNA-Seq technology and developing biological standards with (1) a predictable and finite sequence space in single cells, (2) known gene expression levels among genes and between allelic variants in single cells, and (3) ubiquitous presence across a wide range of tissues. Our aim is to define and sequence up to 250 functional modules comprising multiple gene expression markers based on mouse RNA-Seq data. To benchmark performance, we will target a set of "coexpressed" genes (e.g., VDJ segments from the B-cell IgH loci, X-chromosome inactivation) and use biological allelic exclusion to score sensitivity and specificity in single cells across the reconstructed tissue space. This allows us to tailor methods, train classifiers, and observe deviations from expectation owing to strong biological controls. Later, we will collaborate with other labs in the human cell atlas community to benchmark the performance of our technology. Finally, autoimmunity and inheritable gene expression traits are relevant for many clinical disorders. Therefore, we expect our technology to have a broad impact on clinical or translational medicine, including cancer.

PUBLICATIONS

Lee JH. 2017. De novo gene expression reconstruction in space. *Trends Mol Med* **23:** 583–593.

Lee JH. 2017. Quantitative approaches for investigating the spatial context of gene expression. *Wiley Interdiscip Rev Syst Biology Medicine* **9:** e1369.

In Press

Chen X, Sun Y-CC, Church GM, Lee JH, Zador AM. 2018. Efficient in situ barcode sequencing using padlock probe–based BaristaSeq. *Nucleic Acids Res* **46:** e22.

Lee JH. 2018. Tracing single-cell histories. *Science* **359:** 521–522.

GENETIC AND EPIGENETIC REGULATION OF CANCER AND OTHER HUMAN DISEASES

A. Mills L. Banks S. Balinth C. Ballon Y. Chang M. Fisher A. Jaganathan D. Johnson P. Shrestha S. Sun C. Wu

Our group is focused on determining the genetic/epigenetic basis of cancer and neurodevelopmental syndromes. We have discovered genes impacting these conditions, determined how the encoded proteins work in normal cells, and determined how their dysregulation contributes to disease. These findings have had a major impact and affected how clinicians analyze and treat patients with these syndromes.

Major Discoveries

- Identifying *p63* as a gene affecting development, aging, and cancer
- Defining the genetic basis of autism
- Discovering *CHD5* as a gene that prevents cancer
- Determining that Chd5 loss causes male infertility

p63 in Development, Aging, and Cancer

We discovered *p63*, a gene related to *p53*—a gene encoding a tumor suppressor defective in more than half of all human cancers. Although p63 is similar to p53, its function was not at all clear. We found p63 deficiency leads to premature aging, as p63 is needed for stem-cell renewal. Indeed, lack of p63 causes curvature of the spine, hair loss, and severe skin lesions. Yet, there is a fine balance, as an excess of one version of p63 (ΔNp63α) causes carcinoma—the most prevalent type of human cancer. In contrast, we found that a different version of p63 (TAp63) prevents cancer. Our work showing that TAp63 inhibits tumor growth, even when p53 is absent, was surprising: It had previously been assumed that p53 was needed to prevent cancer. We showed that TAp63 can do the job alone.

We discovered that p63 was needed for development: Its loss causes malformations of the limbs, skin, and palate. This finding led others to interrogate *p63* and reveal that its mutation causes seven different human syndromes involving birth defects affecting the limbs, skin, and palate. By generating mouse models for one of these syndromes—ectrodactyly, ectodermal dysplasia, clefting (EEC) syndrome—we found the reason why some children with EEC syndrome have symptoms that are very severe and even life threatening, whereas other children with EEC—even those in the same family with the same *p63* mutation—have symptoms that are barely noticeable. Within the past year we published our collaborative studies implicating p63 in limb and cartilage development (Kawata et al. 2017; Taniguchi et al. 2017). We are currently working to understand how p63 regulates stem-cell biology and how its perturbation leads to cancer.

CHD5, A New Tumor Suppressor

We discovered *CHD5* as a tumor suppressor mapping to human 1p36—a region of our genomes frequently deleted in cancer. 1p36 deletions occur in many different types of human tumors, including those of the epithelia, brain, and blood. Although this suggested that a cancer-suppressing gene resided in this region, the gene responsible was unknown. By generating mice with deletions and duplications of the genomic region corresponding to 1p36 using chromosome engineering technology—a strategy with which we can generate precise chromosome rearrangements in the mouse—we identified a region of the genome with potent tumor-suppressive activity. Using genetic and molecular approaches, we discovered *CHD5* as the tumor suppressor gene in the region and found that its product turns on a network of tumor suppressors. In addition, we found that *CHD5* is frequently deleted in human glioma. Chromosome engineering proved such a powerful way to identify cancer genes that we also used it for studying neurodevelopmental syndromes, including schizophrenia and autism. This technology was essential for several collaborative studies focused on copy number variations responsible for autism.

We continue to focus on defining the role of CHD5 in chromatin dynamics and deciphering how dysregulation of CHD5 and the pathways it regulates leads to disease. We found that Chd5 uses its plant homeodomains to bind histone 3, and this is essential for tumor suppression. Our work paved the way for further discoveries, and *CHD5* is now known to be mutated in human cancers of the breast, ovary, and prostate, as well as in melanoma, glioma, and neuroblastoma. Furthermore, recent reports indicate that patients with high levels of *CHD5* have much better overall survival than those with low levels. We found that Chd5 is essential for packaging DNA, and that loss of Chd5 leads to improperly packaged DNA that is prone to DNA damage. Intriguingly, Chd5's absence is particularly important during the process of sperm maturation—an event in which the DNA is first unpackaged and then repackaged using an elaborate series of steps—and that deficiency of Chd5 causes male infertility. We discovered that Chd5 is expressed highly in neurons, and that Chd5 plays a pivotal role in the brain, suggesting that inappropriate DNA packaging contributes to neurodevelopmental syndromes such as autism. During the past year, we reported that like CHD5, other members of the CHD family regulate chromatin and are implicated in cancer. We are currently delving deeper into the mechanisms whereby Chd5-mediated regulation of chromatin affects gene expression cascades regulating neuronal stem cells and how dysregulation of this process sets the stage for neurodevelopmental syndromes and cancer.

PUBLICATIONS

Kawata M, Taniguchi Y, Mori D, Yano F, Ohba S, Chung UI, Shimogori T, Mills AA, Tanaka S, Saito T. 2017. Different regulation of limb development by p63 transcript variants. *PLoS ONE* **12:** e0174122.

Mills AA. 2017. The chromodomain helicase DNA-binding chromatin remodelers: Family traits that protect from and promote cancer. *Cold Spring Harb Perspect Med* **7:** a026450.

Taniguchi Y, Kawata M, Ho Chang S, Mori D, Okada K, Kobayashi H, Sugita S, Hosaka Y, Inui H, Taketomi S, et al. 2017. Regulation of chondrocyte survival in mouse articular cartilage by p63. *Arthritis Rheumatol* **69:** 598–609.

CANCER AND HUMAN GENETICS

M. Wigler J. Alexander, P. Andrews, I. Hakker, J. Kendall, S. Li, B. Ma, A. Moffitt, M. Riggs, L. Rodgers, M. Ronemus, J. Rosenbaum, M. Spector, A. Stepansky, Z. Wang, M. Wigler, B. Yamrom, C. Ye, C. Yoon

Our lab works in three areas: genomics, cancer, and human genetics, the latter with a strong emphasis on autism. We do this in close collaboration with three other CSHL laboratories headed by Dan Levy, Alex Krasnitz, and Ivan Iossifov. Our autism sequence data comes from collaboration with the New York Genome Center, where a good portion of sequencing and computational processing also take place.

GENOMICS

Most of our efforts in this area surround issues of structural variation (SV) and genome assembly.

Copy Number

We have a long-term interest in detecting large-scale copy number variation in both normal and cancer genomes and its role in genetic disorders. This interest has led to methods for assessing cancer outcome, detecting cancer cells, identifying oncogenes and tumor suppressors, and exploring the involvement of copy number variation in human genetic disorders. Those applications are discussed later, but recently we published a method called SMASH (Wang et al., *Genome Res 26:* 844 [2016]), which detects copy number variation as cheaply as is theoretically possible by DNA sequencing. We break subject DNA into pieces just large enough for mapping to the genome, ligate them into read-length fragments, sequence those fragments, map the pieces of each read to the genome, and create copy number profiles using specially designed algorithms based on map density. Cost reduction comes from having multiple independent mappings per read, up to six per read fragment.

SVs

Most of our exploration of human genetics derives from analysis of whole-genome sequence data from short-read platforms. SVs, including indels, translocations, inversions, microsatellite instability, pseudogene formation, and transpositions, are a rich source of genetic anomalies. No available genome analysis tool addressed all these things, so we have developed our own. We call it MUMdex, and the underlying approach is to use exact matches (MUMs) to a reference genome. When the distance or orientation between two MUMs in a read differs from their expected distance or orientation in the reference genome, then we have detected a candidate structural variant. Gathering the data over populations, we can then determine whether a structural variant is rare or common. Collecting the data over a family, we gather if a structural variant is de novo in a child. Comparing a cancer to the host genome, we find the somatic structural variants of the cancer. Unlike other methods, MUMdex is indifferent to the type of structural variant or its size. A preprint is published in bioRxiv (Andrews et al., bioRvix doi: 10.1101/078261 [2016]). The tool compares well with the various other more standard tools that are pieced together to do various aspects of the same thing.

Large-Scale Haplophasing

A persistent problem in genomics is haplophasing—how to assign variants in an individual to the proper parent in the absence of the parental DNA. We have devised a solution to the large-scale problem, which we call "HaHa" or haplotyping by halving. Nuclei from the individual are cleaved by cryotome and the resulting subnuclear fragments are sorted by flow sorting. We then use single nuclear sequencing, a methodology pioneered by this laboratory. With

algorithms designed by Dan Levy, using techniques akin to hidden Markov processing, variants in the same subnuclear fraction are aggregated into parental haplophases. The method has been validated in cases in which family DNAs were available and for genomes that were partially phased by long-range sequencers. This work was a collaboration with Partha Mitra.

Assembly by Template Mutagenesis

Another persistent problem in genomics is the genome assembly problem that arises when the structures to be assembled either have a duplicated structure and/or insufficient variation between haplotypes to allow phased assembly with a given sequence platform. One solution to this problem has been the development of long-read sequence platforms such as PacBio and Nanopore. We have approached it by increasing the utility of short-read sequencers, which are vastly cheaper and more accurate than the long-read sequencers. We published the theory for this in 2014 (Levy and Wigler, *Proc Natl Acad Sci 111:* E4632 [2014]). The principle is to introduce random mutations into the template structures, thereby introducing enough variation into each template to make them unique, which in turn permits the long-range assembly of each template from short-read data. We have now reduced the method to practice using partial bisulfite mutagenesis, as published in *Nucleic Acids Research* (Kumar et al. 2018).

CANCER

Most of our effort in the cancer area focuses on risk assessment, response to therapy, and early detection. Although the economic incentives for improvements in clinical care for these advances are poor, the benefits to the patient are potentially huge.

Single-Cell Analysis of Prostate Cancer Biopsy

Many more patients are biopsied for prostate cancer than are treated, and the guides for treatment are quite complex and not standard but are based on morphological Gleason score of core prostate biopsies. We seek to bring genomic methods to this problem. Sparse DNA sequencing of single-cell nuclei from prostate core biopsies is a rich source of quantitative parameters for evaluating neoplastic growth and aggressiveness. These include the presence of clonal populations, phylogenetic structure of those populations, degree of complexity of copy number changes in those populations, and measures of the proportion of cells with clonal copy number signatures. The parameters all show good correlation to the Gleason score, derived from individual prostate biopsy tissue cores and radical prostatectomy surgical specimens. Indeed, these genomic parameters correlate better with the Gleason score of the radical prostatectomy than does the Gleason score of the core biopsies. This is highly relevant because primary treatment decisions are dependent on the biopsy and not the surgical specimen. Thus, single-cell analysis has the potential to augment traditional core histopathology, improving both the standardization and accuracy of risk assessment and hence treatment decisions. Future emphasis will be on making single-cell analysis cheaper so that our methods can become affordable clinical tools. This work was a collaboration with two clinicians, Ashutosh Tewari (Mount Sinai School of Medicine) and Herbert Lepor (New York University Medical School), and was recently published (Alexander et al. 2017).

Measuring Minimal Residual Disease (MRD)

The notion of using molecular techniques to assess MRD has its roots in leukemia and lymphoma, and in particular following chronic myelogenous leukemia (CML) through the presence of the Philadelphia chromosome translocation. In principle, measuring MRD allows the clinician to determine whether therapy is successful and to make a determination whether to continue, switch the therapeutic regimen, or take a radically different approach. In fact, for leukemias and lymphomas, this is common practice. It is now apparent that even solid cancers shed cells and DNA into the blood, and this opens up the possibility of following such cancers in a similar manner. Following the sequence analysis of the presenting neoplasm, we determine the cancer-specific nucleotide mutations. We then apply techniques we developed for error-free sequencing. These allow the detection of single-nucleotide cancer variants at less than one part per million. The basis for this method is template tagging with randomly generated oligonucleotides that we call

"varietal tags" (Hicks et al., U.S. Patent [2012]). We are applying the method to measuring MRD from leukemia and breast cancer in a clinical collaboration with Steve Allen, Jonathan Kolitz and Dan Budman from the Northwell Health hospital system.

Early Detection of Cancer Incidence

Cancer kills by spreading to distant sites. At the time of the first clinical presentation, metastasis has typically already occurred. Were it otherwise, most cancers would be curable by surgery. It follows that there may be a window of time when detection of cancer and its timely extirpation will result in a cure. Some, perhaps most, cancer spreads to distant sites through blood. We plan to analyze the feasibility of the early detection of cancer in blood using genomic analysis. As any early detection method must not cause false alarms, and must be verifiable and actionable, we further propose isolating the suspect cells for further analysis. Our method rests on detecting the presence of recurrent patterns of copy number profile in some of the candidate cells using sparse single-cell sequencing. We show feasibility by simulation, using all copy number profiles from nearly 3900 cases described in The Cancer Genome Atlas and by generating single-cell data in silico. For the latter we devise procedures for sampling from actual diploid single-cell data to produce data for a cell with a given profile. The algorithm for detection requires searching for connected components of highly correlated cell profiles. We recently published our theoretical analysis (Krasnitz et al. 2017).

AUTISM

Autism spectrum disorders (ASDs), a collection of developmental delay syndromes characterized by deficient social skills and communication, receives a strong contribution from genetics. Whereas in the past our focus was genetic contribution from de novo mutation in candidate genes, we now look more broadly into the spectrum of causation.

Shared Ancestral Variation

Given the proportion of autism that is multiplex, we do not expect that all autism is explainable by de novo mutation and that transmission of risk variants plays a role. There has been persistent evidence of shared variation from case–control studies. These approaches universally use a liability threshold model developed by Wang et al. (*Genome Res 26*: 844 [2016]), applied initially to estimate the genetic contribution to quantitative traits such as height on a simple random sample from the population. We have not been satisfied that this method has been correctly applied, and so investigated the question with our own data and methods. We developed a method of analysis (A2DS) that tests generally if shared genomic variants contribute to a disorder (Ye et al. 2017). Using a standard measure of genetic relation, test individuals are compared with a cohort of discordant sib-pairs (CDS) to derive a comparative similarity score. We ask if a test individual is more similar to an unrelated affected than to the unrelated unaffected sibling from the CDS, and then sum over such individuals and pairs. Statistical significance is judged by randomly permuting the affected status in the CDS. In the analysis of published genotype data from the SSC (Simons Simplex Collection) and the AGRE (Autism Genetic Research Exchange) cohorts of children with ASD we find strong statistical significance that the affected are more similar to the affected than to the unaffected of the CDS (p-val ~ 0.00001). Fathers in multiplex families have marginally greater similarity (p-val = 0.02) to unrelated affecters. These results do not depend on ethnic matching or gender. This was a collaboration with Kenny Ye from Albert Einstein College of Medicine.

De Novo Genetic Damage and Loss of Motor Skills

In individuals with ASDs, de novo mutations have previously been shown to be significantly correlated with lower IQ, but not with the core characteristics of ASD: deficits in social communication and interaction and restricted interests and repetitive patterns of behavior. We extend these findings by showing in the Simons Simplex Collection that damaging de novo mutations in ASD individuals are also significantly and convincingly correlated with measures of impaired motor skills. This correlation is not explained by a correlation between IQ and motor skills. We find that IQ and motor skills are distinctly associated with damaging mutations and, in particular, that motor skills are a more

sensitive indicator of mutational severity, as judged by the type and its gene target. We use this finding to propose a combined classification of phenotypic severity: mild (little impairment of both), moderate (impairment mainly to motor skills), and severe (impairment of both). This work has been recently published (Buja et al. 2017), a collaboration with Andreas Buja and Abba Krieger of the University of Pennsylvania.

Contribution from De Novo Indels in the Introns of Candidate Target Genes

Copy number profiling and whole-exome sequencing have allowed us to make remarkable progress in our understanding of the genetics of autism over the past 10 years, but there are major aspects of the genetics that are unresolved. Through whole-genome sequencing, additional types of genetic variants can be observed. These variants are so abundant that knowing which are functional is challenging. We have analyzed whole-genome sequencing data from 510 of the Simons Simplex Collection's quad families and focused our attention on intronic variants. Within the introns of 546 high-quality autism target genes, we identified 63 de novo indels in the affected and only 37 in the unaffected siblings. The difference of 26 events is significantly larger than expected (p-val = 0.01), and using reasonable extrapolation shows that de novo intronic indels can contribute to at least 10% of simplex autism. The significance increases if we restrict to the half of the autism targets that are intolerant to damaging variants in the normal human population—the half we expect to be even more enriched for autism genes. For these 273 targets, we observe 43 and 20 events in affected and unaffected siblings, respectively (p-value of 0.005), and a drop from 26 to 23. A preprint of this work has been recently published (Munoz et al. 2017).

PUBLICATIONS

Alexander J, Kendall J, McIndoo J, Rodgers L, Aboukhalil R, Levy D, Stepansky A, Sun G, Riggs M, Cox H, et al. 2017. Utility of single cell genomics in diagnostic evaluation of prostate cancer. *Cancer Res* **78:** 348–358.

Buja A, Volfovsky N, Krieger A, Lord C, Lash AE, Wigler M, Iossifov I. 2017. Damaging de novo mutations diminish motor skills in children on the autism spectrum. *Proc Natl Acad Sci* **15:** E1859–E1866.

Krasnitz A, Kendall J, Alexander J, Levy D, Wigler M. 2017. Early detection of cancer in blood using single-cell analysis: A proposal. *Trends Mol Med* **23:** 594–603.

Munoz A, Yamrom B, Lee Y-A, Marks S, Lin K-T, Wang Z, Krainer AR, Darnell RB, Wigler M, Iossifov I. 2017. De novo indels within introns contribute to ASD incidence. bioRxiv doi: 10.1101/137471.

Ye K, Iossifov I, Levy D, Buja A, Krieger A, Wigler M. 2017. Measuring shared variants in cohorts of discordant siblings with applications to autism. *Proc Natl Acad Sci* **114:** 7073–7076.

In Press

Kumar V, Rosenbaum J, Wang, Z, Forcier T, Ronemus M, Wigler M, Levy D. 2018. Partial bisulfite conversion for unique template sequencing. *Nucl Acids Res* **46:** e10.

CANCER: SIGNAL TRANSDUCTION

Mikala Egeblad and colleagues study tumors and, in particular, the contributions of the microenvironment in which the cancer cells arise and live. Solid tumors are abnormally organized tissues that contain not only cancer cells, but also various stromal cell types and the extracellular matrix, and these latter components constitute the microenvironment. Communications between the different components of the tumor influence its growth, response to therapy, and ability to metastasize. Among the tumor-associated stromal cells, the lab's main focus is on myeloid-derived immune cells, a diverse group of cells that can enhance angiogenesis and metastasis and suppress the cytotoxic immune response against tumors. Egeblad is interested in how different types of myeloid cells are recruited to tumors and how their behaviors—for example, their physical interactions with cancer cells and other immune cells—influence cancer progression, including metastasis. The Egeblad lab studies the importance of the myeloid cells using mouse models of breast and pancreatic cancer and real-time imaging of cells in tumors in live mice. This enables them to follow the behaviors of and the interactions between cancer and myeloid cells in tumors during progression or treatment. This technique was instrumental when the lab recently showed that cancer drug therapy can be boosted by altering components of the tumor microenvironments, specifically reducing either matrix metalloproteinases (enzymes secreted by myeloid cells) or chemokine receptors (signal receptors on myeloid cells). This year, the Egeblad lab collaborated with Scott Powers' group to understand how normal cells surrounding a tumor promote cancer growth. They found that normal cells signal to tumors through multiple pathways, and that blocking these signals together has the greatest effect on inhibiting tumor growth—offering a new strategy to fight cancer.

The **Douglas Fearon** laboratory studies the interaction between cancer and the immune system. Our underlying premise is that the tumor microenvironment is immune suppressive because cancer cells elicit responses characteristic of wound healing and tissue regeneration. This approach has led to the finding that activated fibroblasts in the tumor stroma mediate immune suppression in several mouse models of cancer, including the autochthonous model of pancreatic ductal adenocarcinoma of the Tuveson lab. Our understanding of the basis of immune suppression is evolving, but we know that it involves the production of the chemokine, CXCL12, by the fibroblastic stromal cells, binding of this CXCL12 by pancreatic cancer cells, and exclusion of T cells from the vicinity of the cancer cells. T-cell exclusion, which also occurs in several types of human adenocarcinomas, causes antagonists of T-cell checkpoints to be ineffective, despite the presence of cancer-specific $CD8^+$ T cells. This immune suppression is interrupted by administering AMD3100, an inhibitor of CXCR4, the receptor for CXCL12, which leads to the rapid accumulation of T cells among cancer cells, thereby uncovering the efficacy of anti-PD-L1 and eliminating cancer cells. Because human pancreatic cancer has certain immunological characteristics of the mouse model, a Phase 1 clinical trial of AMD3100 in patients with pancreatic cancer will be initiated in 2015. Some of our next steps are to determine the biological process that causes cancer cells to express nonmutated, shared antigens and the means by which dormant metastases escape immune elimination.

Darryl Pappin's lab develops chemical and computational methods for analysis of proteins and peptides. These are fundamental tools for proteomics and are vital in many fields of biological investigation. Proteins and peptides are typically analyzed via mass spectrometry, a method that involves fragmenting samples by colliding them with gas atoms in a vacuum. Masses of the resulting fragments are measured, and computer algorithms match the results with molecules whose

amino acid sequences are either known or inferred. Pappin has developed search engines for mass spectrometry data that enable investigators to sift through hundreds of thousands of experimental spectra at a time for database matches. He also seeks to reduce sample complexity via an approach he calls chemical sorting. This includes the use of chelation to enrich phosphopeptides from the total peptide pool and specific affinity-tagged small-molecule inhibitors to segregate classes of kinases or phosphatases for more specific mass spectroscopic analysis.

Despite their large variety of genetic abnormalities, cancer cells have been found to be extremely sensitive to the reversal of certain mutations. **Raffaella Sordella** and colleagues study why cells in certain cancers are responsive to the inhibition of one particular gene or gene product. Why, for instance, do non-small-cell lung cancer (NSCLC) cells that have a particular mutation in the epidermal growth factor receptor (EGFR) respond dramatically to its inhibition by the drug Tarceva? This occurs in 15%–20% of patients, the great majority of whom, within 1–3 years, develop resistance. Various mutations have been implicated in about half of resistant patients. Sordella and colleagues have discovered a new resistance mechanism in a subpopulation of NSCLC cells that are intrinsically resistant to Tarceva. These tumor cells were observed to secrete elevated amounts of a growth factor called transforming growth factor-β (TGF-β), which in turn increases secretion of interleukin-6 (IL-6), an immune-signaling molecule. Significantly, these effects were independent of the EGFR pathway. The team therefore hypothesizes that inflammation is one of the factors that can render a tumor cell resistant to treatment with Tarceva. In other work, Sordella collaborates with the Krainer lab to study whether alternative splicing has a role in the failure of p53-mediated senescence to halt oncogenesis in certain lung cancers.

Nicholas Tonks and colleagues study a family of enzymes called protein tyrosine phosphatases, or PTPs, which remove phosphate groups from proteins and other signaling molecules, such as lipids, in cells. Disruption of PTP function is a cause of major human diseases, and several of the PTPs are potential therapeutic targets for such diseases. Tonks' group seeks to fully characterize the PTP family, understanding how PTP activity is controlled and how PTPs modify signaling pathways. In addition, they are working to determine how those pathways are abrogated in serious illnesses, including cancer, diabetes, and Parkinson's disease. The overall goal is to identify new targets and strategies for therapeutic intervention in human disease. Tonks and colleagues have defined new roles for PTPs in regulating signaling events in breast cancer, identifying three PTPs as novel potential tumor suppressors. They have characterized the regulation of PTP1B by reversible oxidation, showing that it is regulated by covalent modification of the active site by hydrogen sulfide (H_2S) under conditions of ER stress that are linked to protein-folding-related pathologies, such as Parkinson's and Alzheimer's. In addition, they have generated recombinant antibodies that selectively recognize the oxidized conformation of PTP1B; these antibodies display the ability to promote insulin signaling in cells and suggest novel approaches to therapy for diabetes. Finally, they have also discovered a novel mechanism for allosteric regulation of PTP1B activity, offering the possibility of developing small-molecule drugs that could inhibit the phosphatase and thereby modulate signaling by insulin and the oncoprotein tyrosine kinase HER2, potentially offering new ways to treat insulin resistance in type-2 diabetes and breast cancer.

Lloyd Trotman's recent research path begins at his discovery some years ago that the loss of a single copy of a master tumor-suppressing gene called *PTEN* is sufficient to permit tumors to develop in animal models of prostate cancer. His team later found that complete loss of *PTEN* paradoxically triggers senescence, an arrested state that delays or blocks cancer development in affected cells. These findings explained why many patients only display partial loss of this tumor suppressor when diagnosed with prostate cancer. Now the team is researching ways to restore the PTEN protein levels in these patients. This therapeutic approach could slow disease progression and thus greatly reduce the need for surgical removal of the prostate or similar drastic interventions that

carry the risks of incontinence and impotence. Their second approach to combat prostate cancer is to model the lethal metastatic disease in genetically engineered mice. They are developing a novel approach that allows for quick generation and visualization of metastatic disease. The efficacy of existing and novel late-stage therapies, such as antihormonal therapy, can then be tested and optimized in these animals. At the same time, the Trotman lab is exploring the genome alterations associated with metastatic disease and resistance to therapy. To this end, they use single- and multicell genome sequencing techniques developed at CSHL by Drs. Wigler and Hicks.

David Tuveson's lab uses mouse and human tissue models of neoplasia to explore the fundamental biology of these diseases and thereby identify new diagnostic and treatment strategies. His team's main focus is pancreatic cancer, a lethal malignancy that has eluded clinical solutions despite intensive study. The lab's approaches at CSHL run the gamut from designing new model systems of disease to inventing innovative therapeutic and diagnostic platforms for rapid evaluation in preclinical and clinical settings. For example, they have adopted a new method of culturing tissue fragments indefinitely in cell culture, enabling deep analysis with genetic and pharmacological probes. In addition, therapeutic experiments in mouse models have revealed an important role of redox metabolism and stromal interactions on influencing therapeutic response. This year, they used the mouse model system to identify the mechanism of a promising drug treatment for pancreatic cancer. The lab found that using the drug in combination with more standard chemotherapeutic drugs stopped the tumor growth and lengthened life span for the mice, suggesting that the combination therapy may help overcome the drug resistance that is so commonly found in cancers. Tuveson's lab also has a strong link to clinical trials locally and internationally, with confirmation in early-phase trials the ultimate goal. Collectively, their strategy in the preclinical and clinical arenas is codified as the "Cancer Therapeutics Initiative," and this initiative will provide these same approaches to the entire CSHL cancer community.

Dr. Tuveson serves as Director of the Lustgarten Foundation Pancreatic Research Laboratory at CSHL and Director of Research for the Lustgarten Foundation.

Linda Van Aelst's lab studies how aberrations in intracellular signaling involving enzymes called small GTPases can result in disease. They are particularly interested in Ras and Rho GTPases, which help control cellular growth, differentiation, and morphogenesis. Alterations affecting Ras and Rho functions are involved in cancer and various neurodevelopmental disorders. Van Aelst's team has extended its prior study of mutations in a Rho-linked gene called oligophrenin-1 (OPHN1), part of an effort to connect the genetic abnormalities associated with mental retardation to biological processes that establish and modify the function of neuronal circuits. In addition to a role for OPHN1 in activity-driven glutamatergic synapse development, lab members have obtained evidence that OPHN1 has a critical role in mediating mGluR-LTD (long-term depression), a form of long-term synaptic plasticity, in CA1 hippocampal neurons. Their findings provide fresh insight not only into the mechanism and function of mGluR-dependent LTD, but also into the cellular basis by which mutations in OPHN1 could contribute to the cognitive deficits observed in patients. Defects in cortical neurogenesis have been associated with cerebral malformations and disorders of cortical organization. The Van Aelst team discovered that interfering with the function of the Rho activator DOCK7 in neuronal progenitors in embryonic cerebral cortices results in an increase in the number of proliferating neuronal progenitors and defects in the genesis of neurons. In an extension of these studies, the Van Aelst team this year showed that DOCK7 has a central regulatory role in the process that determines how and when a radial glial cell progenitor "decides" to either proliferate (i.e., make more progenitor cells like itself) or give rise to cells that will mature, or "differentiate," into pyramidal neurons. These lines of research provide novel insight into mechanisms that coordinate the maintenance of the neural progenitor pool and neurogenesis.

Hongwu Zheng's lab aims to define the complex biology of malignant glioma pathogenesis, with the ultimate goal of translating the developed knowledge into patient benefits. Although eerily similar in terms of their self-renewal capacity and distinct phenotypic plasticity, malignant glioma cells conspicuously lack the terminal differentiation traits possessed by their normal counterparts—neural progenitors. With the use of multiple approaches combining human cancer genomics, animal modeling, and stem cell biology, Zheng has unraveled the causal relationship between aberrant differentiation and ensuing gliomagenesis. Perhaps more importantly, his team has shown that forced restoration of differentiation capacity within glioma cells can drastically attenuate their tumorigenic potential. This finding fits well with the team's overall strategy, which is to target differentiation control pathways as a novel avenue for malignant glioma treatment. To this end, they have sought to (1) develop various animal models to recapitulate human glioma pathogenesis and use them to trace and investigate in vivo tumor initiation/progression and (2) identify key pathways/players controlling normal and neoplastic neural progenitor cell renewal and fate determination.

THE TUMOR MICROENVIRONMENT AND CANCER PROGRESSION: HOW HOST COMPONENTS HELP TUMORS SPREAD OR CONSTRAIN THEIR PROGRESSION

M. Egeblad J. Albrengues E. Bružas J. Curtis J. Dassler Plenker X. He V. Kuettner L. Maiorino D. Ng L. Puckett M. Shields L. Sun

Solid tumors are aberrant tissues. Like organs, solid tumors are composed of cancer cells and stroma. The stroma is the supportive framework of the organs and includes the extracellular matrix (ECM) as well as immune cells, fibroblasts, adipocytes, and cells of the vascular system. Interactions between epithelium and stroma are essential for normal organ development as well as for tumor progression. In solid tumors, the stroma is also known as the tumor microenvironment.

We study how the tumor microenvironment influences cancer cells in the context of tumor initiation, growth, drug resistance, and metastasis. We use mouse models of breast and pancreatic cancer together with real-time spinning disk confocal and multiphoton microscopy in living mice (known as intravital imaging). This allows us to study how cancer cell proliferation, migration, and survival are influenced by stromal components in real time.

Neutrophil Extracellular Traps Produced during Inflammation Awaken Dormant Cancer Cells

J. Albrengues, M. Shields, D. Ng, V. Kuettner, E. Bružas, L. Maiorino, J. Dassler Plenker

Most cancer patients die not from their original tumor, but from cancer that recurs after metastasizing to a different tissue. Preventing recurrence is therefore a major opportunity to increase cancer patient survival. Recurrence often occurs after a prolonged period of dormancy—a stage in which residual cancer is present but clinically undetectable. Disseminated cancer cells can remain dormant for years, even decades, before recurring, or "awakening," as metastatic cancer. T cells and natural killer cells can eliminate disseminated cancer cells as they start proliferating, preventing them from reaching clinically detectable levels. However, little is known about the cues that drive metastatic recurrence after a long period of clinically undetectable, dormant cancer.

In breast cancer survivors, elevated plasma levels of C-reactive protein, a nonspecific marker of chronic inflammation, are associated with reduced disease-free survival. Smoking is a well established inducer of chronic lung inflammation, and two recent large pooled analysis studies showed that current smoking or prior heavy smoking was significantly associated with an elevated risk of breast cancer recurrence and death from breast cancer. Whether smoking can cause metastasis through induction of inflammation is not clear, but in a mouse model, experimental lung metastasis was previously shown to be increased twofold by tobacco smoke exposure.

Neutrophils are one of the major types of inflammatory cells. In an experimental model of inflammation, it was recently shown that neutrophils are a critical cell type in awakening, but how neutrophils awaken cancer remains unclear. We previously showed that neutrophils can promote metastasis through the formation of neutrophil extracellular traps (NETs)—scaffolds of chromatin with associated cytotoxic enzymes and proteases that are released into the extracellular space where the physiological function is to trap microorganisms. NETs are generated through a signaling process that involves chromatin decondensation and disintegration of the nuclear membrane. Contents from the neutrophil's secretory granules—including neutrophil elastase (NE), cathepsin G (CG), and matrix metalloproteinase 9 (MMP9)—associate with the decondensed chromatin. Finally, the plasma membrane ruptures, and the protease-associated chromatin fibers are released into the extracellular space.

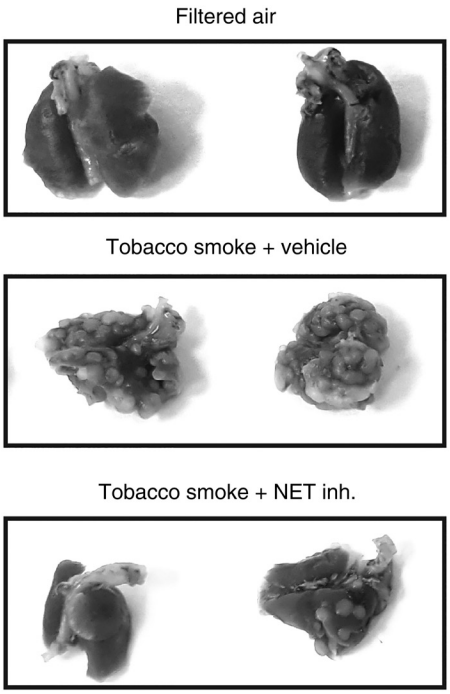


Figure 1. Neutrophil extracellular traps (NETs) promote dormant cancer cell awakening after lung inflammation. Targeting NETs reduced awakening after smoking-induced inflammation. Mice with dormant cancer were housed in filtered air or exposed to tobacco smoke in the presence or absence of an inhibitor of NET formation. Shown are representative photomicrographs of lungs 21 days after initiation of smoking.

We found that sustained experimental lung inflammation caused by tobacco smoke exposure or nasal instillation of lipopolysaccharide (LPS) converted disseminated, dormant cancer cells to aggressively growing metastases in mice. To observe how inflammation influenced the dormant cancer cells, we used our confocal intravital lung imaging (CILI) protocol. After induction of inflammation, we observed NET-like structures around the dormant cancer cells, and this was associated with initiation of proliferation. Excitingly, treatment with NET-digesting, DNase I-coated nanoparticles significantly reduced awakening of the cancer after lung inflammation. Thus, NETs were required for awakening. To determine how NETs induced awakening, we developed an in vitro assay. We found that NETs promoted awakening by enabling two NET-associated proteases, NE and MMP9, to sequentially cleave laminin. We propose that the DNA of the NETs concentrates the neutrophil proteases at their substrate. This is likely critical because only sequential proteolytic remodeling of laminin, first by NE and then MMP9, generated a laminin epitope that induced awakening.

We next determined that NET-cleaved laminin contained an integrin-activating epitope that induced proliferation of dormant cancer cells. This knowledge allowed us to generate blocking antibodies against NET-remodeled laminin, and these antibodies prevented tobacco smoke exposure and LPS from inducing awakening in mice. Preventing dormant cancer cells from awakening is a major opportunity to prolong patient survival, and our data implicate NETs as critical mediators of inflammation-induced metastatic cancer recurrence.

Chronic inflammation and smoking are risk factors in metastatic recurrence. Inflammation is a complex process involving dramatic changes in cellular composition, ECM remodeling, and cytokine production. However, we have now determined that neutrophils recruited during inflammation are key initiators of awakening of dormant cancer, and neutrophils act through the formation of NETs. Our findings set the stage for epidemiological studies testing the possible links among inflammation/smoking, NETs, and recurrence after dormancy in human patients. If such links can be established, we envision that NETs and their downstream effectors could be targeted to reduce the risk of cancer recurrence.

Cancer Cell Chemokine Receptor CCR2 Orchestrates Suppression from the Adaptive Immune Response

X. He, L. Puckett, J. Curtis

The immune system is very efficient at eliminating pathogens that can cause harm to the organism. The immune system also has the potential to eliminate neoplastic cells. The concept of "immune surveillance" was first described more than 50 years ago and refers to the ability of immune cells to detect tumor cells and destroy them. T cells, part of the adaptive immune system, are critical for tumor immune surveillance. Immune surveillance may lead to a period in which cancer cells are kept in check by the immune system, and the tumor neither expands nor regresses. Eventually, tumors develop means to escape immune control. Tumors have multiple mechanisms of escaping immune control, including cancer cell–intrinsic

changes that alter how the cancer cell is recognized by the immune system and extrinsic changes that suppress immune cell activities. As examples of intrinsic changes, cancer cells can decrease the surface expression of major histocompatibility complex (MHC) class I, making them effectively invisible to T lymphocytes. Another mechanism is increased expression of programmed cell death ligand 1 (PD-L1) on the cancer cells. PD-L1 binds the PD-1 receptor on activated T cells, leading to protection against T-cell-mediated killing. Current immunotherapies that target this so-called "immune checkpoint" have led to long-lasting regression in several cancers. Extrinsic mechanisms of immune escape include the down-regulation of costimulatory molecules (e.g., CD86) on antigen-presenting cells; the secretion of cytokines that directly inhibit cytotoxic T lymphocytes (CTLs); and the promotion of regulatory T-cell infiltration. In contrast, infiltration of $CD103^+$ dendritic cells (DCs) has emerged as a mechanism by which tumors may be kept under immune control: $CD103^+$ DCs are highly efficient at acquiring and processing exogenous antigens, and they present the antigens on MHC class I molecules directly to $CD8^+$ CTLs.

Chemokines, or chemotactic cytokines, have critical roles in mediating recruitment of immune cells to sites of inflammation and tumors. For example, the C-C chemokine ligand 2 (CCL2) recruits CC chemokine receptor (CCR2)-expressing immune cells to tumors. The primary role of CCR2 in cancer has, therefore, been considered to be the regulation of immune cell infiltration, and we previously showed that CCL2-mediated recruitment of CCR2-expressing monocytes to tumors after treatment with chemotherapy and these newly recruited monocytes inhibited the chemotherapy response. CCL2/CCR2-mediated recruitment of $CCR2^+$ inflammatory monocytes to the lung has also been shown to promote breast cancer extravasation and metastasis in mice. Furthermore, elevated levels of CCL2 in tumors and serum are associated with advanced disease and poor prognosis in breast carcinoma patients. These findings have sparked interest in targeting the CCR2 pathway for therapeutic benefit in breast cancer.

However, it is not just immune cells that express CCR2: Breast cancer cells also express CCR2. Indeed, CCR2 is up-regulated on breast cancer cells compared with normal breast epithelium. In vitro, cancer cell expression of CCR2 has been proposed to promote tumor growth through increased motility and invasion, cell proliferation, and cell survival. In vivo, the potential role(s) of CCR2 signaling in cancer cells have, however, not been well studied, largely because they were thought to be minor compared with the roles of CCR2 in myeloid cells.

To test the function of CCR2 in breast cancer cells, we used orthotopic transplantation of MMTV-PyMT breast cancer cells as our breast cancer mouse model. We found that *Ccr2* deletion in cancer cells led to reduced tumor growth and twofold longer survival. The longer survival was accompanied by multiple alterations associated with better immune control: increased infiltration and activation of CTLs and $CD103^+$ cross-presenting DCs, as well as up-regulation of MHC class I and down-regulation of checkpoint regulator PD-L1 on the cancer cells. Pharmacological inhibition of CCR2 increased cancer cell sensitivity to CTLs and enabled the cancer cells to induce DC maturation toward the $CD103^+$ subtype. The combination of all these changes likely results in more effective immune surveillance of the $Ccr2^{-/-}$ cancer cells and reduced growth of tumors derived from these cells. Indeed, the $Ccr^{2-/-}$ cancer cells were not growth restricted in $Batf3^{-/-}$ mice lacking the $CD103^+$ DC subtype or in nude mice, lacking CTLs. Thus, breast cancer CCR2 plays a central role in inhibiting immune surveillance.

Our results establish a novel role for CCR2 signaling in cancer cells in orchestration of the suppression of the immune response. These new data, together with our previous findings regarding the role of CCR2 in recruitment of monocytes that protect tumors against chemotherapy, makes CCR2 an attractive target in combination with both chemotherapy and immunotherapy.

Lysyl Oxidases Inhibit Metastasis of Pancreatic Cancer

M. Shields, L. Sun, J. Albrengues [in collaboration with V.M. Weaver, UCSF]

Tumors are often discovered as nodules that are harder than the surrounding tissue. This reflects a stiffening in the ECM and changes in the ECM architecture. The interstitial ECM consists of macromolecules such as fibrillar collagens, fibronectin, and proteoglycans. Type I collagen is the major fibrillar collagen in many tissues, and it forms a scaffold that provides stability.

Type I collagen also has signaling functions mediated by, for example, integrins. The synthesis as well as the proteolytic remodeling of the fibrillar type I collagen increases in many tumors, including breast and pancreatic tumors. Collagen cross-linking in mouse models of mammary carcinoma delays tumor onset and slows tumor progression. Furthermore, collagen architectural structure becomes abnormal with progression of solid tumors: Whereas collagen fibers are curly and oriented in parallel to normal or hyperplastic epithelium, there is a progressive change in the fibers so that they are straighter and mostly perpendicular to the tumor border in the late stages. Lysyl oxidases are a family of five enzymes—lysyl oxidase (LOX) and lysyl oxidase-like (LOXL) 1-4—that can cross-link collagen, thereby generating linearized fibers. LOXs have been shown to promote cancer progression and metastasis in mouse models of breast cancer by generating linearized collagen fibers and enhancing integrin signaling. Consistently, inhibiting LOXs reduces tumor progression and metastasis in several mouse models of breast cancer.

Pancreatic cancer has a very high incidence of metastasis: About 80% of patients present with metastasis and most of the remaining patients develop metastasis within two years of diagnosis. Pancreatic cancer also has a very pronounced tumor microenvironment characterized by deposition of type I collagen. Recent reports have shown that reducing the stromal response, either by inhibiting paracrine hedgehog signaling or genetically ablating collagen-producing myofibroblasts, resulted in more aggressive, undifferentiated tumors. Nevertheless, it is not well understood how the collagen-rich ECM affects pancreatic cancer progression. Gene expression analyses indicate that high expression of LOX family members correlates with worse outcomes in breast and pancreatic cancer patients. Despite these findings, a recent clinical trial in pancreatic cancers showed no survival benefit when chemotherapy was combined with targeting of LOXL2 using a blocking antibody.

LOXs are recognized for their ability to cross-link ECM proteins, but it has largely been forgotten that LOX was first described as an "antioncogene of *ras*," and its expression was shown to inhibit Ras-mediated transformation and tumor growth in NIH 3T3 fibroblasts. Ras mutations are very rare in breast cancer, but they are present in >90% of pancreatic tumors. Collagen content is also different, with pancreatic tumors generally having a more collagen-rich tumor microenvironment than breast cancer.

We set out to test whether prometastatic effects of LOXs on ECM cross-linking or antitumorigenic effects mediated by inhibition of Ras signaling would prevail in orthotopic pancreatic tumor models driven by oncogenic Ras. We found that Lox and Loxl2, the two most highly expressed LOXs, had tumor-suppressive functions in pancreatic cancer, independent of collagen cross-linking. Using intravital imaging, we could observe that on either genetic or pharmacological targeting of LOXs, the pancreatic cancer cells increased their invasion along linearized collagen fibers. LOX inhibition significantly increased distant metastasis. We discovered that Lox and Loxl2 inhibition, specifically in cancer cells with mutant Ras, activated focal adhesion kinase (FAK) signaling, and this activation was independent of the effects of LOXs on collagen cross-linking. Importantly, FAK signaling was required for metastasis. If the mutant Ras was down-regulated in the cancer cells, then LOX inhibition no longer had an effect on FAK signaling. These results suggest that the inhibitory effects of LOX on pancreatic cancer metastasis are dependent on signaling from Ras and FAK.

LOXs, through their prometastatic effect, were attractive targets in pancreatic cancer; however, targeting LOXL2 was ineffective in a clinical trial. Our preclinical work confirms that targeting of LOXs will not benefit pancreatic cancer patients because of the activation of the proinvasive FAK-signaling pathway leading to increased invasion and metastasis. Our results therefore strongly caution against inhibiting LOXs in cancers driven by mutant Ras.

PUBLICATIONS

Albrengues J, Wysocki RW, Maiorino L, Egeblad M. 2017. Preview: Re-cyclin' cell-cycle components to make NETs. *Dev Cell* **43:** 379–380.

Öhlund D, Handly-Santana A, Biffi G, Elyada E, Almeida AS, Ponz-Sarvise M, Corbo V, Oni TE, Hearn SA, Jung Lee E, et al. 2017. Distinct populations of inflammatory fibroblasts and myofibroblasts in pancreatic cancer. *J Exp Med* **214:** 579–596.

Perurena N, Zandueta C, Martínez-Canarias S, Moreno H, Vicent S, Almeida AS, Guruceaga E, Gomis RR, Santisteban M, Egeblad M, et al. 2017. EPCR promotes breast cancer progression by altering SPOCK1/Testican 1-mediated 3D growth. *J Hematol Oncol* **10:** 23.

Roe JS, Hwang CI, Somerville TDD, Milazzo JP, Lee EJ, Da Silva B, Maiorino L, Tiriac H, Young CM, Miyabayashi K, et al. 2017. Enhancer reprogramming promotes pancreatic cancer metastasis.. *Cell* **170:** 875–888.

RESOLVING THE CHEMOKINE/CHEMOKINE RECEPTOR CHALLENGE OF POORLY IMMUNOGENIC TUMORS

D.T. Fearon J. Li Z. Wang M. Yao
A. Pommier R. Yan

The major contribution of the Fearon lab has been a study that addresses the problem of latent metastases in patients with pancreatic ductal adenocarcinoma (PDA). Patients who have had their primary PDA surgically resected often develop metastatic disease despite intraoperative examination of the liver confirming the absence of macrometastatic lesions. These observations lead to the conclusion that latent metastases, detectable only microscopically, were present in these patients and were responsible for the postoperative development of metastatic disease. Latent metastases having a potential for outgrowth had been considered to represent lesions in which cancer cell proliferation is balanced by immune-mediated cancer cell death, but a more recent explanation invokes quiescent, single disseminated cancer cells (DCCs). Single, nonreplicating DCCs have been observed in several cancer types, but whether quiescence is enforced by the microenvironment or is cancer cell autonomous is not known. Immunity, both innate and adaptive, also is likely to have a role in the selection and/or maintenance of latent DCCs, as it has long been suspected based on donor-derived cancer occurring in immunesuppressed recipients of allografts. However, there is an unexplained paradox of immunity preventing the outgrowth of latent metastases while not eliminating latent metastases.

We have studied the metastatic process in the context of an ongoing adaptive immune response because of the occurrence of cancer cell–specific immunity in human and mouse PDA. Livers from patients and mice with PDA contained single DCCs with an unusual phenotype of being cytokeratin-19 $(CK19)^-$ and major histocompatibility complex class I $(MHCI)^-$. We created a mouse model to determine how DCCs develop, their relationship to metastatic latency, and the role of immunity. Intraportal injection of immunogenic PDA cells into preimmunized mice seeded livers only with single, nonreplicating DCCs lacking MHCI and CK19, whereas naïve recipients of PDA cells had macrometastases. We found that T cells select DCCs by eliminating $MHCI^+$-proliferating cancer cells. Transcriptomic analysis of PDA cells with the DCC phenotype showed an endoplasmic reticulum (ER) stress response. Moreover, DCCs showed a lack of activation of the IRE1 pathway of the unfolded protein response, suggesting that DCCs cannot resolve ER stress. Relieving ER stress pharmacologically with a chemical chaperone or genetically by overexpression of spliced XBP1, in combination with T-cell depletion, stimulated outgrowth of macrometastatic lesions containing PDA cells expressing MHCI and CK19.

We conclude that a PDA-specific adaptive immune response selects DCCs in which the ER stress response accounts for both quiescence and resistance to immune elimination. Accordingly, outgrowth of DCCs to macrometastases requires not only relief from the cancer cell–autonomous ER stress response, but also suppression of systemic immunity. Thus, the ER stress response is the cell-autonomous reaction that enables DCCs to escape immunity and establish latent metastases.

PUBLICATIONS

Balachandran VP, Łuksza M, Zhao JN, Makarov V, Moral JA, Remark R, Herbst B, Askan G, Bhanot U, Senbabaoglu Y, et al. 2017. Identification of unique neoantigen qualities in long-term survivors of pancreatic cancer. *Nature* **551:** 512–516.

Connell CM, Raby S, Beh I, Flint TR, Williams EH, Fearon DT, Jodrell DI, Janowitz T. 2017. Cancer immunotherapy trial registrations increase exponentially but chronic immunosuppressive glucocorticoid therapy may compromise outcomes. *Ann Oncol* **28:** 1678–1679.

Connell CM, Raby SEM, Beh I, Flint TR, Williams EH, Fearon DT, Jodrell DI, Janowitz T. 2018. Cancer immunotherapy trials underutilize immune response monitoring. *Oncologist* **23:** 116–117.

Flint TR, Fearon DT, Janowitz T. 2017. Connecting the metabolic and immune responses to cancer. *Trends Mol Med* **23:** 451–464.

Öhlund D, Handly-Santana A, Biffi G, Elyada E, Almeida AS, Ponz-Sarvise M, Corbo V, Oni TE, Hearn SA, Lee EJ, et al. 2017. Distinct populations of inflammatory fibroblasts and myofibroblasts in pancreatic cancer. *J Exp Med* **214:** 579–596.

In Press

Goncalves MD, Hwang SK, Pauli C, Murphy CJ, Cheng Z, Hopkins BD, Wu D, Loughran RM, Emerling BM, Zhang G, et al. 2018. Fenofibrate prevents skeletal muscle loss in mice with lung cancer. *Proc Natl Acad Sci* **115:** E743–E752.

Pommier A, Anaparthy N, Memos N, Kelley ZL, Gouronnec A, Yan R, Auffray C, Albrengues J, Egeblad M, Iacobuzio-Donahue CA, et al. 2018. Unresolved endoplasmic reticulum stress engenders immune-resistant, latent pancreatic cancer metastases. *Science* **360:** eaao4908.

MASS SPECTROMETRY LABORATORY

D. Pappin A. Makarenko J. Snider
K. Rivera N.S. Turna
C. Shaw J. Wilson

Development of Suspension-Trapping Protocols for Mass Spectrometry Sample Preparation

J. Wilson, K. Rivera, N.S. Turna, A. Makarenko, D. Pappin

Despite recent developments in bottom-up proteomics, the need still exists for a fast, uncomplicated, and robust method for comprehensive sample processing. The suspension trapping method combines the advantage of efficient SDS-based protein extraction with rapid detergent removal, reactor-type protein digestion, and peptide cleanup. Proteins are solubilized in 1%–5% SDS, acidified, and introduced into the suspension trapping tip incorporating a depth filter and hydrophobic compartments filled with a neutral pH methanolic solution. The instantly formed fine protein suspension is trapped in the depth filter stack. The methodology allows efficient capture and processing of protein loads down to the low microgram/submicrogram range. The detergent removal takes ~5 min, and tryptic proteolysis of a cellular lysate is complete in as little as 60 min. The Pappin laboratory has been extending this process for almost all sample types, from whole-cell lysates of cultured cells to brain, liver, and muscle tissue and immunoprecipitated samples of protein complexes. The method has also been successfully integrated with a quantitative iTRAQ labeling workflow.

Impact of Detergents on Membrane Protein Complex Isolation

K. Rivera, D. Pappin [in collaboration with S.H. Lin, MD Anderson Cancer Center, Houston; students of the CSHL Protein Purification Course (2016–2017)]

Detergents play an essential role during the isolation of membrane protein complexes, as inappropriate use of detergents may affect the native fold of the membrane proteins, their binding to antibodies, or interaction with partner proteins. The CSHL Purification Course students used cadherin-11 (Cad11) as an example to examine the impact of detergents on membrane protein complex isolation. They found that mAb 1A5 could immunoprecipitate Cad11 when membranes were solubilized by dodecyl maltoside (DDM) but not octylglucoside, suggesting that octylglucoside interferes with Cad11-mAb 1A5 interaction. Furthermore, they compared the effects of Brij-35, Triton X-100, cholate, CHAPSO, Zwittergent 3-12, Deoxy BIG CHAP, and digitonin on Cad11 solubilization and immunoprecipitation, showing that all detergents except Brij-35 could solubilize Cad11 from the membrane. Upon immunoprecipitation, it was found that β-catenin, a known cadherin-interacting protein, was present in Cad11 immune complex among the detergents tested except Brij-35. However, the association of p120 catenin with Cad11 varied depending on the detergents used. Using isobaric tagging chemistries for relative and absolute quantitation (iTRAQ) to determine the relative levels of proteins in Cad11 immune complexes, it was shown that DDM and Triton X-100 were more efficient than cholate in solubilization and immunoprecipitation of Cad11 and resulted in the identification of both canonical and new candidate Cad11-interacting proteins.

In Press

Lee YC, Bååth JA, Bastle RM, Bhattacharjee S, Cantoria MJ, Dornan M, Gamero-Estevez E, Ford L, Halova L, Kernan J, et al. 2018. Impact of detergents on membrane protein complex isolation. *J Proteome Res* **17:** 348–358.

SIGNAL INTEGRATION: A FRAMEWORK FOR UNDERSTANDING THE EFFICACY OF THERAPEUTICS TARGETING

R. Sordella D. Chaluts M. Miller N. Sachan
A. Gray D. Pal P. Safaric Tepes
N. Korimerla P. Pisterzi N. Shirole

The tremendous advances in genomics we witnessed in the past decades have led to the identification of specific "driver" mutations and the development of targeted small-molecule inhibitors to block their functions that have revolutionized the management and treatment of cancers. There was much excitement when these specific inhibitors were employed in the clinic, as they showed initial remarkable efficacies. In the case of lung cancer, this has been the case with erlotinib, gefitinib, afatinib, osimertinib for mutant epidermal growth factor receptor (EGFR), vemurafenib and dabrafenib for mutant BRAF, and crizotinib, ceritinib, and alectinib for ALK and/or ROS1 gene rearrangements. Yet, unfortunately, their clinical success has been limited by the occurrence of resistance. Ultimately in the majority of the cases, despite initial responses these therapies failed to significantly prolong the lives of individuals with cancer as inevitably acquired resistance to these drugs develops. Many mechanisms of resistance to targeted therapy have been identified by us and others, and they can be generally categorized into three main categories: (1) molecular changes in the driver oncogene as a consequence of the occurrence of secondary mutations, (2) activation of critical signaling pathway(s) in a parallel or downstream fashion, and (3) rewiring of the cellular signaling network to drive pro-survival signaling through a different signaling pathway. In the past years we have identified a novel class of resistance that encompasses histological transformation from one cell lineage, such as epithelial, to another. We dubbed cells that have transited into these different cell states as drug-resistant mesenchymal cells (DMR). This population of tumor cells that never completely responds to therapy likely functions as a transition state culminating eventually in a drug-resistant tumor (acquired resistance). These cells are originated by a stochastic/epigenetic mechanism centered on the methylation and consequent silencing of a specific locus containing the miRNA MiR335. Importantly, they are involved in the acquisition of resistance to not only one specific drug, but are present in all tumors we have analyzed and are contributing to resistance in almost all treatments.

Interestingly, we also found that these cells are characterized by intrinsic defects in their DNA double-strand break (DSB) repair capability. This resulted in an increased accumulation of DNA copy number alterations, genetic diversity of cancer cell populations, and improved adaptability to drug treatment. Because these cells are naturally occurring and can be generated by a stochastic/epigenetic program, our findings suggest that the interconversion between different cell states can promote intratumor genetic heterogeneity, spur the tumor's evolution, and hence increase the tumor's fitness. This is important because it argues that the transition into a drug-tolerant/mesenchymal state could provide a mechanism that allows a small subpopulation of tumor cells to withstand an initial destructive attack of drug to enable their survival until they accumulate more permanent resistance mechanisms.

Understanding the mechanisms involved in the survival of this subpopulation of residual tumor cells is critical for identifying successful cancer therapies. If critical signaling pathways can be targeted in a patient in an up-front manner, the emergence of resistance may be delayed or prevented entirely, offering a promising approach to combat the heterogeneity and adaptiveness of most cancers.

Molecular profiling of DMR cells indicated that in addition to EGFR, these cells also express platelet-derived growth factor receptor (PDGFR).

From a molecular standpoint, receptor tyrosine kinases (RTKs) are characterized by an extracellular binding domain, a single transmembrane domain, and a highly conserved cytoplasmic domain. There are 58 known RTKs. Based on their extracellular domain structure, they are divided into 20 subfamilies.

Even though all the RTKs have similar structure and activate similar downstream signaling, they have very distinct biological functions. For example, although EGFR and PDGFR activate similar downstream signaling pathways such as PI3K, MAPK, and STAT3, EGFR causes more proliferative phenotype, whereas PDGFR causes a migratory phenotype. Expression of either wild-type or mutant EGFR does not produce tumors single-handedly, whereas singly PDGFR resulted in tumors.

To explain these differences, it has been proposed that the specificity of the RTKs could be due to (i) the spatial and temporal availability of the ligand and (ii) the intensity of its binding to the receptor. Although this could be very important, only the ligand availability could not explain some of the differences observed in vitro.

Upon ligand binding, RTKs trans-phosphorylate tyrosine residues localized in their cytoplasmic moiety. These tyrosines work as docking sites for effector molecules. As an additional mechanism explaining RTKs' specificity, it has been suggested that different combinations of effectors could modify the subcellular localization of the RTKs, their time of activation, and the activation of distinct signaling pathways.

It has become clear that signaling pathways are not linear and isolated but instead are organized into highly complex and dynamic signaling networks. In this context, highly conserved signaling nodes act as core processors receiving multiple signaling inputs and consolidating them to provide defined, contextual cellular responses such as proliferation, growth, differentiation, etc. Changes in signaling nodes profoundly affect not only the flow of information across the signaling networks, but also the integration of inputs and the biological activity of cells.

We showed that PDGFR, in contrast to EGFR, can directly phosphorylate the suppressor of cytokine signaling 3 (SOCS3), a key negative regulator of the IL6 signaling pathway, at a specific tyrosine residue (Y165). As a consequence, SOCS3 is rapidly degraded and STAT3 activation switched from an oscillatory pattern of activation to a steady one. This leads to distinct cellular behavior and the acquisition of pro-metastatic features. Hence, our findings suggest that SOCS3 phosphorylation by PDGFR could modify the output of the IL-6/STAT3 axis and the drive of therapeutic resistance.

PROTEIN TYROSINE PHOSPHATASES AND THE CONTROL OF SIGNAL TRANSDUCTION

N.K. Tonks C. Bonham, G. Fan, C. Felice, Q. Hu, N. Krishnan, Z. Qian, O. Shrestha, D. Song, P. Venkataramani, S. Zhang, I. Zubiete Franco

As cells encounter stimuli, such as growth factors, cytokines, and hormones, receptors on the cell surface modulate the activities of protein kinases and phosphatases. The functions of these enzymes, which promote the addition and removal of phosphate groups, are coordinated in signal transduction pathways to mediate the cellular response to environmental stimuli. These pathways are of fundamental importance to control of cell function, and their disruption frequently underlies major human diseases. Consequently, the ability to modulate such signal transduction pathways selectively with drugs holds enormous therapeutic potential. In the area of tyrosine phosphorylation-dependent signal transduction, drug discovery efforts to date have emphasized the protein tyrosine kinases (PTKs). Although there have been spectacular successes, challenges remain, including the acquisition of drug resistance. Considering the reversibility of protein tyrosine phosphorylation, there is the potential to manipulate signal transduction pathways at the level of both PTKs and protein tyrosine phosphatases (PTPs). Although the PTPs have been garnering attention as potential therapeutic targets, they remain largely an untapped resource. The long-term objectives of the work of the Tonks lab are to characterize the structure, modes of regulation, and physiological function of members of the PTP family of enzymes. Through basic research to understand the mechanism of action and function of PTPs, the Tonks lab is trying to devise creative new approaches to exploit these enzymes as targets for therapeutic intervention in major human diseases, including diabetes, obesity, and cancer.

During the last year, there was quite a bit of turnover in the lab. Nava Krishnan left to take up a position as Senior Scientist at AbbVie, Gaofeng Fan moved to a faculty position at Shanghai Tech, and Om Shrestha joined Novartis. We were joined by Prabha Venkataramani and Imanol Zubiete Franco as postdocs, Zhe (Changer) Qian as a graduate student from the Stony Brook MCB program, and Qingting Hu as research technician.

Examination of the Function of PTPN23 in Cancer

This year we published our mechanistic analysis of the tumor suppressor function of PTPN23 in models of breast cancer. Although PTPN23 itself may not be a therapeutic target in this context, our analysis illustrates how an understanding of its tumor suppressor function suggests new therapeutic targets and strategies. The *PTPN23* gene is located on chromosome 3p21, a region that is spontaneously lost in 8%–10% of breast cancers, and low expression of PTPN23 coincides with poor survival. Our previous work identified PTPN23 as an important regulator of mammary epithelial cell migration and invasion. Now, in an orthotopic transplantation model, we observed that when PTPN23-deficient Comma 1Dβ cells were transplanted in the cleared mammary fat pad, suppression of PTPN23 induced severe primary tumor development in 52 weeks (18 out of 24) and a few cases (two out of 24) of lung metastasis. Immunohistochemistry analysis detected extensive phosphorylation of Tyr 142 in β-catenin in tumor samples, but not in mammary glands repopulated with control cells or normal mammary glands. We have shown that the primary effects of PTPN23 in this context are exerted through FYN, a SRC family kinase. Under normal conditions, PTPN23 dephosphorylates the autophosphorylation site in FYN, thereby suppressing its activity. Following suppression of PTPN23, FYN activity is elevated,

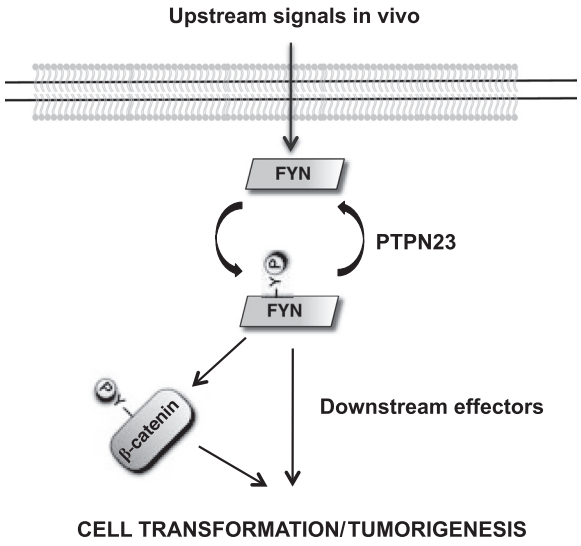


Figure 1. Model to illustrate our proposed mechanism for breast tumorigenesis induced by suppression of PTPN23 in vivo. In this working model, the suppression of PTPN23 impaired the dephosphorylation of FYN on Tyr420 in cancer cells. Enhanced autophosphorylation of the tyrosine in the activation loop of FYN promoted kinase activity, which contributed to cell transformation in vivo, involving aberrant phosphorylation of substrates, such as β-catenin, and disrupted regulation of downstream signal cascades. (Reproduced from Zhang et al. 2017.)

leading to enhanced phosphorylation of substrates, including Tyr 142 in β-catenin (Fig. 1). More recently, we turned to tumor xenograft models because we could generate data in 8–10 weeks, compared with ~52 weeks in the transplantation model. We used RNAi to suppress PTPN23 in BT474 cells and tested the effects of AZD0530 (saracatinib), a small-molecule inhibitor of SRC family kinases, on tumor xenografts in SCID/Beige mice. Whereas suppression of PTPN23 accelerated tumor formation from BT474 xenografts, treatment with AZD0530 reversed the effect, coincident with inhibition of FYN activation. In a complementary study, we tested the effects of using CRISPR-Cas9-based gene targeting to suppress PTPN23 alone or together with FYN in CAL51 breast cancer cells. Using a xenograft model in SCID/Beige mice, we observed that suppression of FYN rescued tumor outgrowth due to loss of PTPN23 in CAL51 cells, consistent with the effects of AZD0530. These data provide mechanistic insights into the tumor suppressor function of PTPN23 and suggest that FYN may be a therapeutic target for tumors defined by loss of heterozygosity of PTPN23.

Previous studies have shown that PTPN23 is also an ESCRT (endosomal sorting complexes required for transport)-associated protein. ESCRTs are multimeric protein complexes mediating a number of important physiological processes, including multivesicular body (MVB) formation, cytokinetic abscission, autophagy, membrane repair, and retroviral budding. In a loss-of-function CRISPR-Cas9 screen, we showed that PTPN23 was required for mouse AML cell proliferation and survival, but dispensable for normal mouse bone marrow cells. Validation in three additional human AML cells showed similar dependency on PTPN23. Interestingly, this reflects a positive role for PTPN23 in AML cell proliferation and survival, which contrasts with its tumor suppressor function in breast cancer. We constructed a series of PTPN23 truncations and found that the minimal segment of human PTPN23 that can rescue the single-guide mouse knockout is a fragment comprising residues 1–872, which does not contain the PTP domain. Currently, we are testing whether PTPN23 functions through ESCRT pathways in this context and identifying binding partners of PTPN23 to define their contribution to its function in supporting AML cell proliferation and survival.

Characterizing Allosteric Inhibitors of PTP1B and Their Impact on Various Signaling Pathways to Exploit Further the Potential of this Phosphatase as a Therapeutic Target

We have continued to focus on novel approaches to exploiting PTP1B as a therapeutic target in various human diseases. PTP1B plays a positive role in promoting signaling events downstream from the oncoprotein tyrosine kinase HER2; consequently, small-molecule inhibitors of PTP1B may represent novel therapeutics for treatment of tumorigenesis and malignancy associated with elevated HER2. Previously, we showed that Trodusquemine/MSI-1436 is a novel allosteric inhibitor of PTP1B, which attenuates HER2-dependent tumorigenesis and abrogates metastasis in the NDL2 mouse model of breast cancer. This led to MSI-1436 being tested in a Phase 1 clinical trial in metastatic breast cancer patients (see ClinicalTrials.gov: NCT02524951), in collaboration with the Northwell Health Monter Cancer Center. Our current focus remains on examining the effect of MSI-1436 on tyrosine phosphorylation in HER2-positive breast cancer cells, such as BT474, to define its mechanism of action. In particular, our goal is to define how MSI-1436 may

be able to overcome the de novo and acquired resistance to Herceptin (trastuzumab) that is observed in HER2-positive cancer patients.

In addition, we are continuing to investigate the role of PTP1B as a regulator of TRKB function and as a therapeutic target for Rett syndrome (RTT). We have observed that Trodusquemine/MSI-1436 ameliorates the effects of MECP2 loss in RTT mouse models, similar to the effects of a structurally and mechanistically distinct, active site-directed inhibitor, CPT-157633. In collaboration with David Katz (Case Western) we are testing the effects of our PTP1B inhibitors in a distinct mouse model of RTT to allow us to examine more broadly those aspects of the RTT phenotype that are affected by PTP1B, with the goal of defining end points that would be used in a clinical trial in RTT patients.

Although Trodusquemine shows efficacy in an injectable format, like many PTP1B inhibitors identified to date it also has limited oral bioavailability. In an exciting development, we have identified DPM-1001—an orally bioavailable analogue of MSI-1436 that has taken us in an unanticipated direction. Our mechanistic studies have now revealed that this molecule binds copper with high affinity ($k_D = 5$ nM) and unique specificity, which enhanced its potency as a PTP1B inhibitor. DPM-1001 displayed antidiabetic properties that were associated with enhanced signaling through insulin and leptin receptors in animal models of diet-induced obesity, consistent with its ability to target PTP1B. More recently, we have explored further its ability to chelate copper specifically.

The levels of copper, an essential element in living organisms, are under tight homeostatic control. Inactivating mutations in ATP7B, a P-type Cu-ATPase that functions in copper excretion, promote aberrant accumulation of the metal, primarily in the liver and brain. This condition underlies Wilson's disease, a severe autosomal recessive disorder characterized by profound hepatic and neurological deficits. Current treatment regimens rely on the use of broad specificity metal chelators as "decoppering" agents; however, there are side effects that limit their effectiveness. Focusing on the ability of DPM-1001 to chelate copper, we have shown that treatment of cell models, including fibroblasts derived from Wilson's disease patients, eliminated adverse effects associated with copper accumulation. Furthermore, treatment of the "toxic milk" mouse model of Wilson's disease with DPM-1001 lowered the levels of copper in liver and brain, removing excess copper by excretion in the feces, while ameliorating symptoms associated with the disease. These data suggest that DPM-1001 may form the basis for a new therapeutic approach to Wilson's disease.

Recently, there have been reports suggesting the potential application of copper suppression as a cancer therapy, and we are excited about the potential applications of our discovery to cancer. In particular, the reports that copper exerts a stimulatory effect on MEK, which is a critical element in the RAS-MAPK signaling pathway that has been implicated in many cancers, has stimulated considerable interest in clinical trials with copper chelators such as tetrathiomolybdate. In fact, Linda Vahdat (Weill Cornell Medical College) led a trial of tetrathiomolybdate in advanced triple-negative breast cancer patients with encouraging results. We have now investigated the mechanism underlying the specificity of DPM-1001 for copper and identified an analogue, DPM-1003, that has the same chemical composition as DPM-1001, but in which copper chelation is markedly impaired. We have now tested these compounds in a panel of triple-negative breast cancer cells and observed that DPM-1001 was cytotoxic, whereas DPM-1003 was not. Interestingly, we observed that the levels of copper in these cells were elevated compared with controls, and the effects of DPM-1001 coincided with suppression of those levels of copper. Furthermore, we showed that DPM-1001 inhibited tumor growth in mice. We implanted MDA-MB-231 cells in mammary fat pads and waited for large, palpable tumors to form before commencing administration of the compounds. In contrast to treatment with saline or DPM-1003, DPM-1001 abrogated tumor growth. Overall, DPM-1001 offers several advantages over tetrathiomolybdate, and we plan to continue its characterization in cell and animal cancer models. Furthermore, in our attempts to define the importance of copper in tumorigenesis, we have identified several members of the family of protein kinases that bind copper and may be regulated by the metal. Currently, we are testing these potential targets further in a variety of cancer models.

Redox Regulation of PTP Function

PTP1B is a major regulator of the signaling pathways initiated by insulin, which controls glucose uptake

and metabolism, and leptin, which controls appetite. Gene-targeting studies showed that PTP1B-null mice are healthy, display enhanced insulin sensitivity, do not develop type 2 diabetes, and are resistant to obesity when fed with a high-fat diet (HFD). This generated considerable interest in PTP1B as a therapeutic target for treatment of diabetes and obesity. Major programs in the industry focused on developing small-molecule inhibitors of PTP1B have been frustrated by technical challenges arising from the chemical properties of the PTP active site. In particular, although it is possible to generate potent, selective, and reversible active site-directed inhibitors, the tendency for such molecules to be highly charged, such as pTyr-substrate mimetics, presents problems with respect to their oral bioavailability and limits their drug development potential. As a result, the industry views PTP1B, and other PTPs, as challenging. Consequently, innovative strategies are required to generate inhibitors of this highly validated target that may be readily exploited for drug development.

We have taken the approach of trying to harness a physiological mechanism for redox regulation of PTP1B function that reflects a new tier of control of tyrosine phosphorylation–dependent signaling. Previously, we observed that the activity of PTP1B is attenuated by reversible oxidation of an essential cysteinyl residue at the active site of the enzyme. The architecture of the PTP active site is such that this essential cysteinyl residue displays unique properties that favor its role as a nucleophile in catalysis, but also render it prone to oxidation. Insulin stimulation of mammalian cells leads to enhanced and localized production of intracellular H_2O_2, which causes reversible oxidation of PTP1B and inhibition of its enzymatic activity; this, in turn, concomitantly promotes the signaling response to insulin. We have shown that mild oxidation of PTP1B, such as occurs in response to insulin, results in profound conformational changes in the active site of the enzyme that transiently inhibit substrate binding and catalysis. These structural changes are reversible and the enzyme can be reduced back to its active state. Therefore, reversible oxidation of PTP1B in response to insulin provides a mechanism for fine-tuning the signaling response to the hormone. Previously, we used phage display to identify conformation-sensor antibodies, such as scFv45, that recognize the reversibly oxidized form of PTP1B (PTP1B-OX) selectively and stabilize this inactive state, inhibiting its reactivation by reducing agent and thereby inhibiting phosphatase activity. We showed that expression of these antibodies in cells enhanced insulin-induced signal transduction. These data provide proof-of-concept that stabilization of the oxidized, inactive form of PTP1B with appropriate therapeutic molecules may offer a novel paradigm for phosphatase drug development.

To test the hypothesis that the effects of the PTP1B-OX-directed conformation-sensor antibodies could be mimicked by small molecules, we developed an assay that allowed us to conduct a pilot screen of the LOPAC library of compounds (Sigma-Aldrich). This revealed two such hits, one of which was sanguinarine. We screened additional analogs of sanguinarine, ultimately focusing on chelerythrine. Chelerythrine bound directly to PTP1B-OX, at a site that overlapped with the binding site for scFv45, and inhibited its reduction and reactivation; however, it did not affect TCPTP, the closest relative of PTP1B. Treatment of HEK cells with chelerythrine led to enhanced insulin signaling, and this effect was abrogated by expression of catalase, which promotes decomposition of H_2O_2. Furthermore, treatment of HFD-fed C57Bl6/J mice with chelerythrine resulted in weight loss, whereas there was no effect of treatment with saline or protopine, an inactive analog of chelerythrine. Treatment with chelerythrine, but not saline or protopine, also improved glucose tolerance and insulin sensitivity, coincident with enhanced phosphorylation of the insulin receptor β-subunit and AKT in liver. In addition, it led to enhanced leptin signaling in the hypothalamus. These data are consistent with a mechanism in which chelerythrine, like scFv45, stabilizes PTP1B in an inactive, oxidized conformation and enhances insulin and leptin signaling. This work illustrates a novel paradigm for inhibiting the signaling function of PTP1B, which focuses on the critical pool of PTP1B that is responsible for regulation of insulin and leptin signaling—the pool of the enzyme that is acutely regulated following insulin and leptin stimulation. By focusing on this pool of PTP1B, the approach may minimize the potential for complications arising from inhibition of the native enzyme as a whole. Overall, our data suggest that this approach may represent the basis for a novel strategy for therapeutic intervention in diabetes and obesity.

PUBLICATIONS

Gurung P, Fan G, Lukens JR, Vogel P, Tonks NK, Kanneganti, T-D. 2017. Tyrosine kinase SYK licenses MyD88 adaptor protein to instigate IL-1α-mediated inflammatory disease. *Immunity* **46:** 635–648.

Zhang S, Fan G, Hao Y, Hammell M, Wilkinson JE, Tonks NK. 2017. Suppression of protein tyrosine phosphatase N23 predisposes to breast tumorigenesis via activation of FYN kinase. *Genes Dev* **31:** 1939–1957.

In Press

Krishnan N, Konidaris KF, Gasser G, Tonks NK. 2018. A potent, selective and orally bioavailable inhibitor of the protein tyrosine phosphatase PTP1B improves insulin and leptin signaling in animal models. *J Biol Chem* **293:** 1517–1525.

UNDERSTANDING PTEN AND PROSTATE CANCER METASTASIS

L. Trotman A. Ambrico H. Cox G. Matthew J. Vilaro
I. Casanova-Salas M. Lee D. Nowak K. Watrud

Resistance to therapy of metastatic prostate cancer (CRPC) is responsible for the annual deaths of some 250,000 U.S. men. Whereas there is considerable progress in development of improved antihormone therapy for treatment of metastatic disease, this standard-of-care approach will invariably fail at some point.

Our focus is to understand the mechanisms driving human prostate cancer (PC) in its most lethal form: metastatic disease. We have studied the human genetics behind the transition from indolent to lethal metastatic PC and combined it with viral transgene delivery into the prostate. With this approach, we have succeeded in generating a unique, fast, and faithful mouse model for metastatic PC. We have termed this system RapidCaP as it allows us to generate any genetically mutant mouse PC with a much accelerated time frame compared with breeding-based approaches. Now, we use RapidCaP for analysis and therapy of metastatic disease.

At the same time, we aim to better understand how the PTEN tumor suppressor works. This has given us new insights into how cancer successfully attacks PTEN protein in early stage PC before the PTEN gene is mutated.

Selective Killing of PTEN-Deficient Cancer Cells

G. Matthew, A. Naguib, K. Watrud, A. Ambrico, T. Herzka, I. Casanova-Salas, M.F. Lee, W. Zheng [in collaboration with N. El-Amine, D. Pappin, CSHL; E. Di Francesco, J. Marszalek, MD Anderson Cancer Center; C.R. Reczek, Navdeep S Chandel, Northwestern Medical School]

A hallmark of advanced PC is the concomitant loss of PTEN and p53 function. Although there is progress in development of improved drugs for treatment of metastatic disease, the vast majority of patients are in one type of treatment approach: antihormone therapy. This approach goes back to discoveries made in the 1940s and has improved the condition of most men with the disease. However, it is not a cure: Emergence of androgen resistance is invariably observed, even after 70 years of improvements. We are interested in searching for alternatives that specifically target the PTEN-deficient cancer cells.

The genes for p53 and PTEN genes are most frequently inactivated in lethal PC. Their diminished activity provides a growth, proliferative, and antiapoptotic advantage. In mouse PC models, loss of p53 does not result in malignancy, yet codeletion of *Pten* and *Trp53* genes causes prostatic adenocarcinoma, illustrating the need to combat the cooperative power of these combined genetic lesions. To selectively eliminate such cells, we screened cytotoxic compounds on $Pten^{-/-};Trp53^{-/-}$ fibroblasts and their *Pten*-WT reference. Highly selective killing of *Pten*-null cells can be achieved by deguelin, a natural insecticide. Deguelin eliminates *Pten*-deficient cells through inhibition of mitochondrial complex I (CI). Five hundred–fold higher drug doses are needed to obtain the same killing of *Pten*-WT cells, although deguelin blocks their electron transport chain equally well. Selectivity arises because mitochondria of *Pten*-null cells consume ATP through complex V, instead of producing it. The resulting glucose dependency can be exploited to selectively kill *Pten*-null cells with clinically relevant CI inhibitors, especially if they are lipophilic. In vivo, deguelin suppressed disease in our genetically engineered mouse model of metastatic PC. Our data thus introduce a vulnerability for highly selective targeting of incurable PC with inhibitors of CI.

The PHLPP2 Phosphatase Protects MYC and Is a Target for Prevention of PC Progression

D.G. Nowak, K. Watrud, I. Casanova-Salas, A. Ambrico [in collaboration with C-H. Pan, CSHL; K. Cohen Katsenelson, A.C. Newton, University of California, San Diego; John E. Wilkinson, University of Michigan, Ann Arbor]

Metastatic PC commonly presents with targeted, biallelic mutation of PTEN and TP53 tumor suppressor

genes. In contrast, however, most candidate tumor suppressors are part of large recurrent hemizygous deletions, such as the common chromosome 16q-deletion, which involves the AKT-suppressing phosphatase, PHLPP2.

Using RapidCaP, a genetically engineered mouse model of *Pten-Trp53*-mutant metastatic PC, we found that complete loss of *Phlpp2* effectively blocks prostate tumor growth and progression to otherwise lethal metastasis. We show that *Phlpp2* activates Myc, a key driver of PC and metastasis. Mechanistically, *Phlpp2* dephosphorylates the Thr-58 site of Myc, thus directly increasing MYC stability. Finally, we show that small-molecule inhibitors of PHLPP2 can suppress MYC and cause cell death. Our findings reveal how PTEN-deficient tumors can thrive in the absence of AKT activation, driven by PHLPP2 stabilization of MYC. They also suggest that the frequent hemizygous deletions on chromosome 16q present a druggable vulnerability for targeting the MYC protein through PHLPP2 phosphatase inhibitors.

PUBLICATIONS

Chen M, Nowak DG, Narula N, Robinson B, Watrud K, Ambrico A, Herzka TM, Zeeman ME, Minderer M, Zheng W, et al. 2017. The nuclear transport receptor Importin-11 is a tumor suppressor that maintains PTEN protein. *J Cell Biol* **216:** 641–656.

Senturk S, Shirole NH, Nowak DG, Corbo V, Pal D, Vaughan A, Tuveson DA, Trotman LC, Kinney JB, Sordella R. 2017. Rapid and tunable method to temporally control gene editing based on conditional Cas9 stabilization. *Nat Commun* **8:** 14370.

In Press

Alexander J, Kendall J, McIndoo J, Rodgers L, Aboukhalil R, Levy D, Stepansky A, Sun G, Chobardjiev L, Riggs M, et al. 2018. Utility of single-cell genomics in diagnostic evaluation of prostate cancer. *Cancer Res* **78:** 348–358.

CANCER MEDICINE LABORATORY

D. Tuveson	B. Alagesan	E. Elyada	E. Lee	H. Patel	C. Tonelli
	L. Baker	D. Engle	M. Ma	D. Plenker	M. Yao
	G. Biffi	F. Froeling	S. Moran	C. Schoepfer	G. Yordanov
	R. Burkhardt	L. Guerriero	K. Miyabayashi	B. Spielman	K. Yu
	D. Cheng	C-I. Hwang	T. Oni	J. Thalappillil	
	A. Deschenes	R. Koike	Y. Park	H. Tiriac	

Our laboratory investigates the fundamental biology of pancreatic ductal adenocarcinoma (PDA), a disease that is currently the third leading cause of death due to cancer in the United States. We use our basic insights to design new therapeutic and diagnostic strategies to treat and detect PDA. To enable these studies, in addition to using genetically engineered PDA mouse models that we have generated over the past 20 years, we have started to analyze large data sets of sequenced PDA tumors in efforts to find new cancer drivers, revealing a suite of noncoding mutations that likely participate in the pathology of PDA (Feigin et al. 2017). Furthermore, we are also using organoid models of human and mouse PDA, which we published in 2015. These organoid model systems have revealed previously uncharacterized facets of this deadly disease. For example, organoids were used to investigate the process of pancreatic cancer metastasis, revealing epigenetic alterations as a driving event in this process (Roe et al. 2017). Furthermore, we have found using organoid and mouse models that oncogenic Kras promotes selective mitophagy to rewire intermediary metabolism and promote redox homeostasis. Finally, we have generated and characterized a large collection of human PDA organoids and found a striking correlation between organoid therapeutic profiling, transcriptional patterns, and patient outcomes. Clinical trials are now being planned to evaluate the organoids as a new approach to prospectively choosing therapies for patients.

Noncoding Mutations in PDA

This work was done in collaboration with A. Biankin (Glasgow) and M. Schatz (Johns Hopkins).

Although coding mutations have been causally linked to tumor progression, less is understood about the role of noncoding somatic mutations. To investigate this further, we analyzed the whole genome sequences of 308 previously published PDA samples, and compared them with the germline sequences to identify noncoding somatic mutations. We collaborated with the laboratory of Michael Schatz to develop a pipeline to assist in the analysis, termed GECCO (Genomic Enrichment Computational Clustering Operation). GECCO is able to quickly analyze somatic noncoding alterations and identify commonly mutated regulatory regions. We found that genes whose regulatory regions contain recurrent noncoding mutations were enriched in PDA pathways, including axon guidance and cell adhesion, and newly identified processes, including transcription and homeobox genes. Furthermore, we identified mutations in protein-binding sites correlating with differential expression of proximal genes and experimentally validated effects of mutations on expression. Additional bioinformatics approaches yielded an expression modulation score that quantified the strength of gene regulation imposed by each class of regulatory elements, and thereby we found that the strongest elements were most frequently mutated, suggesting a selective advantage. Our detailed single-cancer analysis of noncoding alterations identified regulatory mutations as candidates for diagnostic and prognostic markers, and suggested new mechanisms for tumor evolution (Feigin et al. 2017). These findings have prompted us to perform whole-genome sequencing on the human PDA organoids and matched germline DNA in efforts to establish cell models that can be used to functionally interrogate these noncoding alterations.

Metastasis in Pancreatic Cancer

This work was done in collaboration with M. Egeblad and C. Vakoc (CSHL) and T. Hollingsworth (Nebraska).

Metastasis is a main cause of PDA lethality, and oftentimes this is present during the initial diagnosis of the disease. We initiated a collaborative project with

the laboratory of Chris Vakoc to compare the epigenetic features of neoplastic cells found in the primary tumors with those of metastatic cells cultured as PDA organoid models. We found substantial gene expression differences caused by differential enhancer activation. One subset of altered enhancer activation was delineated by increased H3K27 acetylation, and this correlated closely with the increased expression of the Foxa1 transcription factor. Foxa1 activated a foregut endoderm differentiation pattern in organoids, and was required for metastasis in PDA transplantation models (Roe et al. 2017). Another subset of enhancer alterations was conversely a result of decreased H3K27 acetylation and the corresponding decrease of nearby gene expression, and this correlated with the increased expression of the transcriptional repressor Engrailed 1 (En1). En1 was also shown to be crucial to promote PDA metastasis (C-I Hwang, in prep.). Therefore, a suite of transcription factors promotes PDA metastasis, in part because of altered enhancer activation gene expression changes, which we have termed "epigenetic reprogramming." Current efforts are focused on determining new therapeutic vulnerabilities downstream from epigenetic reprogramming in PDA cells. In addition, we are investigating how the tumor microenvironment sculpts the process of epigenetic reprogramming, and searching for additional processes that promote metastasis.

Mitophagy Promotes Pancreatic Cancer

This work was done in collaboration with G. Evan (Cambridge), J. Girnun (Stony Brook), K. McLeod (Chicago), and K. Vousden (Glasgow).

Activating mutations in KRAS are found in almost all cases of PDA—yet effective clinical targeting of oncogenic KRAS remains elusive. A better understanding of KRAS-dependent PDA-promoting pathways could lead to the identification of cancer-cell-specific vulnerabilities and the development of new treatments. We have found that oncogenic $Kras^{G12D}$ induces BNIP3L/Nix expression and a selective mitophagy program that restricts flux of glucose to the mitochondria and enhances cellular redox control. Genetic ablation of Nix delays progression of PanIN to PDAC and improves survival in a $Kras^{G12D}$-driven murine model of pancreatic cancer in vivo. Although Nix ablation initially leads to an accumulation of mitochondria in vivo, adaptations occur to normalize mitochondrial content to the level of Nix-wild-type mice by the time these mice reach survival endpoint. Taken together, our results reveal the Kras-Nix mitophagy program as a novel mechanism to promote aerobic glycolysis, redox homeostasis, and disease progression in pancreatic cancer (B Alagesan et al., submitted; see Fig. 1). Furthermore, this prompts us to more closely address the metabolic dependencies in PDA.

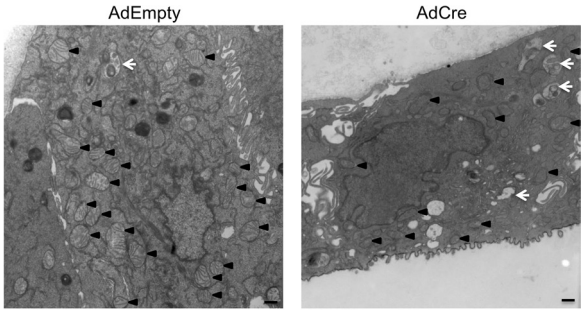


Figure 1. Oncogenic Kras-G12D promotes mitophagy in pancreatic ductal organoids. LSL-Kras-G12D organoid infected with empty-vector adenovirus (*left*) or adeno-cre (*right*). Mitochondria are indicated by arrowheads, and mitophagy is indicated by arrows.

Human Organoid Platform for Development of PDA Therapeutics

This work was done in collaboration with J. Crawford (Hofstra Northwell School of Medicine), A. Krasnitz (CSHL), and E. Li (Stony Brook University).

In collaboration with several hospitals, including Northwell Health, Thomas Jefferson University, Memorial Sloan Kettering Cancer Center (MSKCC), Stony Brook University, Johns Hopkins Medical Institute, and Winthrop Hospital, we have obtained more than 200 human pancreatic tumor tissues and successfully established more than 150 organoids from pancreatic tumor, metastases, and biopsy samples. We have determined gene expression and somatic mutations for each sample, and profiled the response of each organoid to a panel of available and investigational therapies. Our combined assessment has identified collections of patient-derived organoids with selective responses to certain therapies and a correlation with gene expression and the mutational spectrum. Preliminary data show a strong correlation between the results of organoid therapeutic profiling and the responses of patients who were treated in parallel (Tiriac et al. 2018). We are planning to determine the effectiveness

of organoids for the improvement of patient outcomes in clinical studies that are being designed.

PUBLICATIONS

Feigin ME, Garvin T, Bailey P, Waddell N, Chang DK, Kelley DR, Shuai S, Gallinger S, McPherson JD, Grimmond SM, et al. 2017. Recurrent noncoding regulatory mutations in pancreatic ductal adenocarcinoma. *Nat Genet* **49:** 825–833.

Roe JS, Hwang CI, Somerville TDD, Milazzo JP, Lee EJ, Da Silva B, Maiorino L, Tiriac H, Young CM, Miyabayashi K, et al. 2017. Enhancer reprogramming promotes pancreatic cancer metastasis. *Cell* **170:** 875–888.

In Press

Tiriac H, Belleau P, Engle DD, Plenker D, Deschênes A, Somerville TDD, Froeling FEM, Burkhart RA, Denroche RE, Jang GH, et al. 2018. Organoid profiling identifies common responders to chemotherapy in pancreatic cancer. *Cancer Discov* **8:** 1112–1129.

RAS AND RHO GTPASES AND THE CONTROL OF SIGNAL TRANSDUCTION

L. Van Aelst C. Fitzpatrick S. Nakamuta
N.B. Gallo Y. Tai
A. Mirko M. Wang

Research in our laboratory is focused on the role of Ras and Rho GTPase family members in signal transduction. Ras and Rho family members play key roles in cellular activities controlling cell growth, differentiation, and morphogenesis. Alterations in Ras and Rho functions have been implicated in cancer, as well as brain/mental disorders—the latter including intellectual disability (ID), autism, schizophrenia, epilepsy, and mood disorders. Our interests lie in understanding how defects in Ras- and Rho-linked proteins contribute to the development of these diseases/disorders. Toward this end, the lab has continued to define the functions of selected GTPases, their regulators, and their effectors, using animal models of cancer and neurodevelopmental/neurological disorders. Below are highlighted our key projects.

Dual Role for DOCK7 in Driving Tangential Migration of Interneuron Precursors in the Postnatal Forebrain

DOCK7 is a member of the evolutionarily conserved DOCK180-related protein superfamily. The DOCK180 family emerged as a distinct class of Rac and/or Cdc42 GTPase guanine nucleotide exchange factors (GEFs), which have diverse cell type–specific functions. We initially identified DOCK7 as a novel activator of Rac GTPases that is highly expressed in the developing brain. Significantly, sequence variations in DOCK7 have been reported in individuals with epileptic encephalopathy and schizophrenia; however, the function(s) of DOCK7 in neuronal development and/or function have remained largely elusive.

We previously documented that DOCK7 plays a critical role in the genesis of new neurons from radial glial progenitor cells (RGCs) in the embryonic neocortex, and it does so by controlling apically directed interkinetic nuclear migration via its interaction with the centrosome-associated protein TACC3 (transforming acidic coiled-coil-containing protein 3). These findings, combined with our more recent data unveiling the presence of DOCK7 in areas of postnatal/adult neurogenesis, prompted us to explore whether DOCK7 also controls the genesis of new neurons during postnatal/adult stages. Although the bulk of neuronal precursor generation and migration in the mammalian brain occurs during the embryonic period, these processes do persist in restricted areas of the postnatal/adult brain. Among them is the ventricular-subventricular zone (V-SVZ), which in rodents is located along the walls of the brain lateral ventricles. Here, each day, neural stem cells give rise to thousands of interneuron precursors, termed V-SVZ neuroblasts, that migrate tangentially over a long distance to the olfactory bulb (OB), where they differentiate into various subtypes of local circuit interneurons. This continual influx of new neurons enables constant modification of OB neural circuits, a property vital for olfactory information processing.

Significantly, we found that DOCK7 is essential for the tangential migration of V-SVZ neuroblasts to the OB, but not for the generation or proliferation of these cells. Interestingly, using a molecular replacement strategy combined with live-cell imaging, we found that DOCK7 governs the migration of V-SVZ neuroblasts by controlling both LP extension and nucleokinesis. Moreover, we found that DOCK7 controls these two processes by acting on distinct pathways. Specifically, it controls LP stability/growth via a Rac-dependent pathway, likely by stabilizing microtubules (MTs), while it regulates somal/nuclear translocation by modulating F-actin dynamics at the cell rear via a previously unrecognized $p116^{Rip}$ (myosin phosphatase-RhoA–interacting protein)-dependent pathway. Thus, our studies identified DOCK7 as a novel molecule that regulates both phases of V-SVZ neuroblast migration. Thereby, it not only offers new insight into DOCK7 function, but it also provides a

greater understanding of the mechanisms that govern the multifaceted steps undertaken by tangentially migrating V-SVZ neuroblasts in the postnatal forebrain.

Multifunctional Role of the X-linked Intellectual Disability (XLID) Protein Oligophrenin-1 in Cognition and Behavior

Oligophrenin-1 (*OPHN1*), which encodes a Rho-GTPase-activating protein, was the first identified Rho-linked ID gene. It was initially identified by the analysis of a balanced translocation t(X;12) observed in a female patient with mild ID. Subsequent studies revealed the presence of *OPHN1* mutations in families with a syndromic form of ID, with affected individuals commonly showing mild to severe ID and behavioral problems. Most of the *OPHN1* mutations identified to date were shown, or predicted, to result in loss of OPHN1 function; however, the pathophysiological role of *OPHN1* has remained poorly defined. To gain insight into how mutations in *OPHN1* could lead to the cognitive deficits, we initially focused on defining OPHN1's role in hippocampal plasticity associated with learning and memory. Our studies unveiled multiple roles for OPHN1 at hippocampal CA1 synapses. In addition to an essential role in controlling activity-driven glutamatergic synapse development, we found that temporal regulation of OPHN1 translation plays a critical role in mGluR-dependent long-term depression (LTD), a form of plasticity linked to drug addiction and cognitive disorders that is dependent on rapid new protein synthesis. We subsequently showed that OPHN1's involvement in mGluR-LTD and its ability to control glutamatergic synapse development are independent of each other. Whereas the latter requires OPHN1's Rho-GAP activity and association with Homer 1b/c proteins, the former is dependent on OPHN1's interaction with Endophilin A2/3. Together, our findings provided first insights into how mutations in *OPHN1* could contribute to the cognitive deficits in individuals with *OPHN1* mutations.

Apart from the above-described role for OPHN1 in hippocampal plasticity and learning, we recently discovered that it also plays a critical role in moderating stress-induced depressive-like behaviors in a learned helplessness (LH) model of depression. In this model, animals are exposed to an unpredictable and uncontrollable stressor, and subsequently evaluated for their coping capabilities (i.e., resilient vs. helpless/depression-like behavior) in a testing session that consists of an aversive but escapable stressor. We found that *Ophn1*-deficient mice ($Ophn1^{flox//Y}$;β-$actin^{Cre-/+}$) subjected to the LH procedure show a marked increase in helpless behavior, compared with control littermates. Importantly, *Ophn1*-deficient mice did not display altered motor activity and were capable of learning and performing the avoidance task. On further examination of the neuronal cell types involved, we found that *Ophn1* deficiency in parvalbumin (PV), but not somatostatin (SOM), interneurons promotes LH. Moreover, and importantly, we found that *Ophn1* depletion in PV interneurons exclusively in the prelimbic area of the medial prefrontal cortex (mPFC) is sufficient to induce helpless behavior. Together, these findings uncovered the importance of the XLID protein OPHN1 in the establishment of resilience to stress and revealed its involvement in mPFC PV interneuron function.

TGF-β/Smad Signaling through DOCK4 Facilitates Lung Adenocarcinoma Metastasis

The cytokine TGF-β plays an important, albeit complex, role in epithelial tumorigenesis. During early stages of tumorigenesis TGF-β typically functions as a tumor suppressor. At later stages, however, it can act as a potent promoter of multiple events driving the metastatic process, which comprises local motility/invasion, entry of cancer cells into the blood stream (intravasation), exit from the blood vessels (extravasation), and colonization of distant organs. The relevance of TGF-β signaling for disease progression has been particularly recognized in tumors in which cancer cells retain the core TGF-β signaling components, as is frequently the case in breast and lung cancers. However, a major remaining challenge is the identification of TGF-β target genes that drive specific events during metastasis, especially because TGF-β modulates gene expression in a highly cell- and context-specific manner. Although some progress has been made in the context of breast cancer metastasis, the genes and mechanisms that mediate the prometastatic effects of TGF-β in lung adenocarcinoma (ADC) remain largely unknown.

To identify molecular mechanisms that mediate the prometastatic effects of TGF-β in lung ADC, we

took a candidate gene approach and started by scrutinizing members of the DOCK180-related protein superfamily, which, as mentioned above, emerged as a distinct class of Rac and/or Cdc42 GEFs. We found that in lung ADC cells, expression of DOCK4, but not other DOCK180-family members, is rapidly and robustly induced by TGF-β in a Smad-dependent manner. Subsequently, we found that DOCK4 is a direct TGF-β/Smad target gene, and, importantly, that high DOCK4 expression correlates with activated TGF-β signaling and poor prognosis in human lung ADC. These findings prompted us to assess the role of DOCK4 in mediating the prometastatic effects of TGF-β in lung ADC in vivo in a mouse model. Remarkably, we found that DOCK4 induction is essential for TGF-β-driven lung ADC metastasis. Specifically, we observed that blockade of TGF-β-mediated DOCK4 induction attenuates the ability of lung ADC cells to extravasate into distant organ sites, resulting in a marked reduction in metastatic burden. At a cellular level, our evidence supports a model in which TGF-β-induced DOCK4 facilitates extravasation by stimulating lung ADC cell protrusive activity, motility, and invasion, without promoting epithelial-to-mesenchymal transition (EMT), and, intriguingly, that it does so by driving Rac1 activation. So far, Rac1 has only been linked to TGF-β via a noncanonical pathway. Thus, our findings identified the atypical Rac1 activator DOCK4 as a novel key component of the TGF-β/Smad pathway that promotes lung ADC cell extravasation and metastasis.

Recently, we expanded this line of research toward identifying genes that mediate lung ADC organ-specific metastases, with a particular focus on genes that mediate colonization of specific organs. To this end, we established a multiple organ metastasis model system using mice injected intracardially with lung ADC cell populations derived from primary tumors arising in $Kras^{G12D}/p53^{-/-}$ mice. With this model system, we initiated an in vivo RNA interference (RNAi) screen designed to assess the possible involvement of a select set of genes in modifying the potential of lung ADC cells to metastasize to specific host organs. These studies were performed in collaboration with Kenneth Chang. Interestingly, we found that silencing of this gene set resulted in a significant decrease in occurrence of brain metastases and an increase in the occurrence of bone metastases. These data suggest that one or more genes in this set modulate(s) lung ADC metastatic cell homing to and/or colonization of the brain and bone. We are currently investigating which of these genes, individually or in combination, modulate organ-specific homing and/or colonization of lung ADC cells in the brain and bone. These studies have the potential to identify novel mediator(s) and suppressor(s) of lung ADC metastases to brain and bone.

PUBLICATIONS

Nakamuta S, Yang Y-T, Wang C-L, Gallo NB, Yu J-R, Tai Y, Van Aelst, L. 2017. Dual role for DOCK7 in driving tangential migration of interneuron precursors in the postnatal forebrain. *J Cell Biol* **216:** 4313–4330.

In Press

Van Aelst L, Downward J, McCormick F, eds. 2018. *Ras and cancer in the 21st century*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

UNDERSTANDING MALIGNANT GLIOMA PATHOGENESIS

H. Zheng C. Wu L. Xu

Malignant glioma is the most common type of brain tumor and is currently incurable. The lack of effective treatments highlights an urgent need to identify mechanism-based therapeutic approaches. Our group uses an integrated approach by combining glioma genetics, in vivo animal modeling, as well as in vitro biochemical and cellular assays to dissect the molecular programs that potentially govern glioma pathogenesis. Our long-term goal is to unravel the molecular and developmental programs vital for glioma initiation and progression so as to enable the development of targeted therapeutic strategies against this dreadful disease. During the past year, we have continued to focus our research efforts around two major areas: (1) to develop genetically engineered animal model systems to recapitulate the genetic and pathological aspects of various human glioma subtypes and use these glioma animal models as tools to dissect in vivo tumor propagation and their response to experimental therapeutic treatments; and (2) to identify the genetic and epigenetic pathways governing normal and neoplastic neural stem and glioma cell fate determination. Below are some of our highlights.

ATRX Facilitates Neural Differentiation and Its Depletion Promotes Gliomagenesis

Epigenetic regulation controls cell fate determination and organism development. Not surprisingly, its dysregulation potentiates tumor pathogenesis. Previously, we had conducted a project to comprehensively probe epigenetic modulators that might potentially be involved in regulation of neural stem and glioma cell self-renewal or differentiation. In addition to various chromatin regulators known to be important for glioma maintenance, we identified a gene called α-thalassemia/mental retardation syndrome X-linked (ATRX), whose depletion facilitates self-renewal of glioma cells against differentiation induction. ATRX belongs to the chromatin remodeling factor SNF2 family, and its germline mutation causes an X-linked developmental disorder featuring mental retardation, developmental delay, and α-thalassemia. Importantly, frequent loss-of-function mutations of *ATRX* have been identified in a variety of human cancer types, including pediatric and adult malignant gliomas, pancreatic neuroendocrine tumor, neuroblastoma, osteosarcoma, and soft-tissue sarcomas. Although ATRX is functionally known to partner with the histone variant H3.3 chaperone protein DAXX and assist nucleosome assembly at heterochromatic regions independently of the DNA replication process, how ATRX exerts its glioma suppressor function is unclear.

Dynamic nucleosome deposition, eviction, and remodeling are essential for epigenetic regulation, which orchestrates cellular lineage and organism development. Because ATRX is functionally involved in the replication-independent nucleosomal assembly, we reasoned that ATRX might play a pivotal role in facilitating neural lineage differentiation. Indeed, we found that CRISPR-Cas9-induced depletion of ATRX in mouse-derived glioma cells promoted cell proliferation by blocking differentiation. To further test whether ATRX deficiency affects differentiation of an untransformed neural stem cell (NSC), we isolated primary control and ATRX-depleted mouse NSCs from litter-mated animals. After being transduced with a neuronal transcriptional factor NeuroD1 and induced for differentiation, the control NSCs, as expected, differentiated and stopped proliferation; in contrast, the ATRX-deficient NSCs retained their progenitor profiles and kept proliferating. These observations support the notion that ATRX exerts its glioma suppressor function by facilitating the neural differentiation. Consistently, we showed that deletion of *ATRX* markedly enhanced gliomagenic capacity of mouse $p53^{-/-}$;$Pten^{-/-}$ NSCs after being orthotopically grafted into recipient animals. Analysis of NSC lineage differentiation by RNA-sequencing (RNA-Seq) and ChIP-seq further confirmed that ATRX plays a crucial role in controlling neural development. Moreover, our data revealed that depletion of ATRX in mouse NSCs strongly compromised their neuronal

lineage maturation, but it had a limited effect on their proliferation or differentiation into other cell lineages, such as astrocytes and oligodendrocytes—suggesting that the neuronal lineage progenitors might be the cell of origin of the ATRX mutant gliomas. To extend our discovery to the in vivo setting, we are currently conducting lineage tracing in the engineered animal model to investigate the role of ATRX during glioma initiation and progression.

ATRX Deficiency Dictates the Telomere Maintenance Pathway during Human Cell Immortalization

Telomeres are the specialized structures located at the end of each eukaryotic chromosome that function to protect chromosomal stability and genomic integrity. A telomere shortens along with each cell division because of the intrinsic chromosomal end replication problem. Thus, to maintain their long-term replicative capacity, the proliferative eukaryotic cells rely on their ability to counteract the progressive telomere erosion. Failure to do so causes telomere loss, which subsequently leads to chromosome defects, genetic instability, senescence, or cell death. Whereas a majority of human cancers achieve their long-term proliferation by activating telomerase, which adds de novo telomere repeats to chromosome ends, the remaining 10%–15% of them use a recombination-based mechanism, termed alternative lengthening of telomeres (ALT), for their telomere elongation. Importantly, a high concordance has recently been identified between the presence of ALT activity and *ATRX* mutations in many human cancers, including pediatric and adult malignant gliomas, suggesting an intimate connection between ATRX status and ALT pathway activation. But despite the strong genetic evidence, the molecular mechanisms and pathways underpinning *ATRX* loss–associated glioma pathogenesis and their effect on telomere maintenance remain poorly understood.

In collaboration with Dr. Lieberman's group at Wistar Institute, we have revealed that depletion of ATRX in human NSCs and glioma cells induces ALT activation, irrespective of preexisting telomerase activity and telomere length. Mechanistically, we showed that ATRX deficiency triggers telomere destabilization, which not only triggers ALT activation and elongates telomeres, but also activates telomere DNA damage response, which subsequently suppresses mutant cell growth. By using the ATAC-seq (assay for transposase-accessible chromatin [ATAC] with high-throughput sequencing) method to probe chromatin accessibility with hyperactive Tn5 transposase, we further revealed that ATRX is required for maintenance of normal telomeric nucleosomal density and chromatin configuration. Depletion of ATRX causes progressive telomeric chromatin decompaction, which eventually triggers telomere destabilization and ALT activation.

The unlimited proliferation of cancer cells requires a mechanism to prevent telomere shortening. To test whether ATRX deficiency was sufficient to induce ALT and subsequently drive human cell immortalization, we generated CRISPR knockout for *ATRX* in a well known immortalization system—the SV40-large T (LT) transformed but telomerase-negative IMR-90 fibroblasts. In contrast to the *ATRX*-wild-type control fibroblasts that ceased proliferation and entered crisis after a number of cell divisions, we found that ATRX-deficient cells underwent early cell growth arrest as a result of telomere DNA damage responses. More importantly, we found that in contrast to the immortalized cells that emerged from the control cultures, which achieve their immortalization by reactivating the endogenous telomerase activity, the immortalized ATRX-depleted cells were all telomerase-negative and used the ALT pathways for telomere maintenance. These findings suggest that *ATRX* loss dictates telomere maintenance pathway during human cell immortalization and tumor progression.

Together, our study reveals that ATRX loss does not directly activate ALT, but instead disrupts histone variant H3.3-mediated telomere nucleosome assembly, which subsequently triggers telomere DNA replication dysfunction, which eventually leads to telomere destabilization and ALT activation. At the molecular level, we showed that ATRX deficiency triggers progressive telomere decompaction and the increase of nucleosome-free regions, which predate the onset of telomere dysfunction. Importantly, these findings suggest that ALT deficiency not only promotes glioma pathogenesis by compromising the central nervous system cell lineage development but, paradoxically, it also triggers telomere dysfunction, which requires the evolvement of specific mechanism during tumor progression to compensate for its telomere defect. Given that *ATRX* is

highly mutated in both pediatric and adult malignant gliomas, it stands to reason that identification of the critical factors behind the yet unidentified compensatory mechanism will hold high promise in developing efficient therapeutic strategies and drugs against this devastating disease. We are currently focusing our effort on this task, and hope to uncover novel mechanism-based therapeutic targets in the near future.

PUBLICATIONS

Calvert AE, Chalastanis A, Wu Y, Hurley LA, Kouri FM, Bi Y, Kachman M, May JL, Bartom E, Hua Y, et al. 2017. Cancer-associated IDH1 promotes growth and resistance to targeted therapies in the absence of mutation. *Cell Rep* **19:** 1858–1873.

Singh DK, Kollipara RK, Vemireddy V, Yang XL, Sun Y, Regmi N, Klingler S, Hatanpaa KJ, Raisanen J, Cho SK, et al. 2017. Oncogenes activate an autonomous transcriptional regulatory circuit that drives glioblastoma. *Cell Rep* **18:** 961–976.

NEUROSCIENCE

How does the brain encode stimuli from the outside world, within and across sensory modalities, to generate specific perceptions that trigger complex behaviors? How is the brain shaped by sensory experience, and what modifications occur in neuronal circuits that allow us to learn and remember? These are the questions guiding the work of **Florin Albeanu**, who is using the olfactory bulb and olfactory cortex of mice as the subject of his current studies. Airborne chemicals, translated into neuronal signals by specific receptors in the nose, are sent directly to the olfactory bulb. Advances in optical imaging and optogenetics combined with electrophysiological recordings enable Albeanu and colleagues to monitor and/or alter patterns of activity at unprecedented synaptic and millisecond resolution, in real time, as animals are engaged in various behaviors. For survival, rodents need to identify the smells of objects of interest, such as food, mates, and predators, across their recurring appearances in the surroundings despite apparent variations in their features. Furthermore, animals aptly extract relevant information about their environment across different sensory modalities, combining olfactory, visual, or auditory cues. By recording neuronal activity in the input and output layers of the olfactory bulb, as well as feedback from olfactory cortical areas and neuromodulatory signals, Albeanu and his team aim to understand computations the bulb performs and how this information is decoded deeper in the brain. They have recently published evidence suggesting that the mouse olfactory bulb is not merely a relay station between the nose and cortex, as many have supposed. Using optogenetic tools and a novel patterned illumination technique, they discovered that there are many more information output channels leaving the olfactory bulb for the cortex than there are inputs received from the nose. They are currently investigating how this diversity of bulb outputs is generated, as well as how downstream areas, such as the piriform and parietal cortex, make use of such information during behaviors.

The study of decision-making provides a window into the family of brain functions that constitute cognition. It intervenes between perception and action and can link one to the other. Although much is known about sensory processing and motor control, much less is known about the circuitry connecting them. Some of the most interesting circuits are those that make it possible to deliberate among different interpretations of sensory information before making a choice about what to do. **Anne Churchland**'s lab investigates the neural machinery underlying decision-making. Lab members use carefully designed paradigms that encourage experimental subjects to deliberate over incoming sensory evidence before making a decision. Recent results show that rats and humans have a statistically similar decision-making ability. To connect this behavior to its underlying neural circuitry, the researchers measure electrophysiological responses of cortical neurons in rodents as they perform designated tasks. The lab's current focus is on parietal cortex, which appears to be at the midpoint between sensory processing and motor planning. Churchland and colleagues also use theoretical models of varying complexity to further constrain how observed neural responses might drive behavior. This approach generates insights into sensory processing, motor planning, and complex cognitive function.

Hiro Furukawa's lab studies receptor molecules involved in neurotransmission. Its members mainly focus on the structure and function of NMDA (*N*-methyl-D-aspartate) receptors—ion channels that mediate excitatory transmission. Dysfunctional NMDA receptors cause neurological disorders and diseases, including Alzheimer's disease, Parkinson's disease, schizophrenia, depression, and stroke-related ischemic injuries. The Furukawa lab is working to solve the three-dimensional structure of the very large NMDA receptor by dividing it into several domains. They seek to understand the pharmacological specificity of neurotransmitter ligands and allosteric modulators in

different subtypes of NMDA receptors at the molecular level. Toward this end, they use cutting-edge techniques in X-ray crystallography to obtain crystal structures of the NMDA receptor domains and validate structure-based functional hypotheses through a combination of biophysical techniques, including electrophysiology, fluorescence analysis, isothermal titration calorimetry, and analytical centrifugation. Crystal structures of NMDA receptors serve as a blueprint for creating and improving the design of therapeutic compounds with minimal side effects for treating neurological disorders and diseases. During the last several years, the team discovered and mapped several regulatory sites in specific classes of NMDA receptors—progress that now opens the way to the development of a new potential class of drugs to modulate the receptor activity.

Josh Huang and colleagues study the assembly and function of neural circuits in the neocortex of the mouse. The neocortex consists of a constellation of functional areas that form a representational map of the external (sensory, social) and internal (visceral, emotional) world. These areas are strategically interconnected into elaborate information processing networks that guide behavior. The group's overarching hypothesis is that, at the cellular level, cortical processing streams and output channels are mediated by a large set of distinct glutamatergic pyramidal neuron types, and functional neural ensembles are regulated by a diverse set of GABAergic interneuron types. Understanding cortical circuit organization requires comprehensive knowledge of these basic cellular components. The Huang lab uses state-of-the-art genetic approaches to systematically target cell types and facilitate the application of a full set of modern techniques for exploring cortical circuits. Among GABAergic interneurons, the chandelier cell is one of the most distinctive cell types that control pyramidal neuron firing at the axon initial segment. Huang and colleagues are studying the developmental specification, activity-dependent circuit integration, and functional connectivity of chandelier cells—an entry point toward understanding a local circuit module. Regarding pyramidal neurons, they are systematically characterizing the developmental origin, axon projection pattern, and input connectivity of multiple classes of pyramidal neuron types, focusing on the forelimb motor cortex. They combine a range of approaches that include genetic and viral engineering, genetic fate mapping, gene expression profiling, cellular imaging, electrophysiology, and behavior analysis. Recently, they began to integrate their studies in the context of the motor cortex control of forelimb movements.

Adam Kepecs and colleagues are interested in identifying the neurobiological principles underlying cognition and decision-making. They use a reductionist approach, distilling behavioral questions to quantitative behavioral tasks for rats and mice that enable the monitoring and manipulation of neural circuits supporting behavior. Using state-of-the-art electrophysiological techniques, they first seek to establish the neural correlates of behavior and then use molecular and optogenetic manipulations to systematically dissect the underlying neural circuits. Given the complexity of animal behavior and the dynamics of neural networks that produce it, their studies require quantitative analysis and make regular use of computational models. The team also has begun to incorporate human psychophysics to validate its behavioral observations in rodents by linking them with analogous behaviors in human subjects. Currently, the team's research encompasses the study of (1) the neural basis of decision confidence, (2) the division of labor among cell types in prefrontal cortex, (3) how the cholinergic system supports learning and attention, and (4) social decisions that rely on stereotyped circuits. A unifying theme is the use of precisely timed cell-type and pathway-specific perturbations to effect gain and loss of function for specific behavioral abilities. This year, the Kepecs lab was able to link foraging decisions—the choice between staying or going—to a neural circuit and specific cell types in the prefrontal cortex. In other work, they identified a class of inhibitory neurons that specializes in inhibiting other inhibitory neurons in the cerebral cortex and conveys information about rewards and punishment. Through manipulations of genetically and anatomically defined neuronal elements, the team hopes to identify

fundamental principles of neural circuit function that will be useful for developing therapies for diseases such as schizophrenia, Alzheimer's disease, and autism spectrum disorder.

Alexei Koulakov and colleagues are trying to determine the mathematical rules by which the brain assembles itself, with particular focus on the formation of sensory circuits such as those involved in visual perception and olfaction. The visual system of the mouse was chosen for study in part because its components, in neuroanatomical terms, are well understood. What is not known is how projections are generated that lead from the eye through the thalamus and into the visual cortex, how an individual's experience influences the configuration of the network, and what parameters for the process are set by genetic factors. Even less is known about the assembly of the neural net within the mouse olfactory system, which, in the end, enables the individual to distinguish one smell from another with astonishing specificity and remember such distinctions over time. These are among the challenges that engage Koulakov and his team.

Understanding the link between neural circuits and behavior has been the focus of research in **Bo Li**'s lab. The Li lab is particularly interested in studying the synaptic and circuit mechanisms underlying reward processing, attention, and learning and memory, as well as synaptic and circuit dysfunctions responsible for maladaptive behaviors that are related to major mental disorders. The Li lab integrates in vitro and in vivo electrophysiology, imaging, molecular, genetic, optogenetic, and chemogenetic techniques to probe and manipulate the function of specific neural circuits—with a focus on the fear and reward circuits—in the rodent brain, and determine how these circuits participate in adaptive or maladaptive behavioral responses in various tasks.

Partha Mitra is interested in understanding intelligent machines that are products of biological evolution (particularly animal brains) with the basic hypothesis that common underlying principles may govern these "wet" and "dry" intelligent machines that are transforming the present economy. Mitra initiated the idea of brain-wide mesoscale circuit mapping, and his laboratory is involved in performing such mapping in the mouse (http://mouse.brainarchitecture.org) and the marmoset (in collaboration with Japanese and Australian scientists at the RIKEN Brain Science Institute and Monash University).

Mitra spent 10 years as a member of the theory department at Bell Laboratories and holds a visiting professorship at IIT Madras, where he is helping establish the Center for Computational Brain Research. He has an active theoretical research program in machine learning and control theory, where he is using tools from statistical physics to analyze the performance of distributed/networked algorithms in the "thermodynamic" limit of many variables.

Pavel Osten's lab works on identification and analysis of brain regions, neural circuits, and connectivity pathways that are disrupted in genetic mouse models of autism and schizophrenia. Osten hypothesizes that (1) systematic comparison of multiple genetic mouse models will allow determination of overlaps in pathology—neural circuit endophenotypes—responsible for the manifestation of neuropsychiatric disorders, and (2) neural circuit–based classification of autism and schizophrenia will provide key circuit targets for detailed mechanistic studies and therapeutic development. Osten and colleagues have developed the first systematic approach to the study of neural circuits in mouse models of psychiatric diseases, based on a pipeline of anatomical and functional methods for analysis of mouse brain circuits. An important part of this pipeline is high-throughput microscopy for whole-mouse brain imaging, called serial two-photon (STP) tomography. This year, they used this method to describe the first whole-brain activation map representing social behavior in normal mice. They are currently focusing on using this approach to study brain activation changes in two mouse models of autism: the 16p11.2 df/+ mouse model, which shows an increased propensity to seizures and hyperactivity, and the CNTNAP2 knockout mouse model, which shows abnormal social behavior.

Stephen Shea's lab studies the neural circuitry underlying social communication and decisions. They use natural social communication behavior in mice as a model to understand circuits and processes that are evolutionarily conserved and, therefore, shared broadly across species, likely contributing to disorders such as autism. Shea and colleagues have examined how emotion and arousal enable mice, via their olfactory systems, to store memories of other individuals and related social signals. The team has exploited the intimate relationship between memory and emotion to effectively create memories in anesthetized mice, allowing them unprecedented access to neurobiological processes that typically only occur during behavior. The lab has been making a detailed analysis of the changes in neural connections that underlie odor memory. The team is particularly focused on an enigmatic cell type (granule cells or GCs) that has long been hypothesized to be crucial for memories, but has resisted direct study. They have developed methods for recording, giving them the first glimpse of the dynamics of these cells while the animal is learning an odor. The results show unexpectedly complex population dynamics among the GCs that were independently predicted by a model of odor learning developed in Alexei Koulakov's lab. The two labs are collaborating to discern how GC population activity gets integrated by olfactory bulb output neurons and to pinpoint the synaptic circuit that underlies this form of learning. In parallel, another member of the lab is using imaging techniques to determine how memories are stored among broad neuronal ensembles at a different level of the system. Recently, the lab made a key breakthrough, developing the ability to record from GCs in awake animals and discovering that their activity is dramatically modulated by state of consciousness. Finally, the Shea lab completed a series of studies of a different form of social recognition: auditory recognition of pup vocalizations by their mothers. Through this research, they have shown that a mouse model of Rett syndrome shows deficits in communication and learning not unlike those in human patients. Grants from the Simons and Whitehall foundations are allowing the lab to extend this work by directly linking these deficits to the action of the gene *MeCP2* in the auditory cortex.

The **Jessica Tollkuhn** lab seeks to understand how transient events during brain development exert lasting effects on gene expression, circuit function, and, ultimately, behavior. They study how sex-specific neural circuits in rodents are established and modulated by the gonadal hormones estrogen and testosterone. The cognate receptors for these hormones are nuclear receptor transcription factors, which orchestrate modification of local chromatin environment and, thus, exert long-term effects on gene expression. However, the genes regulated by these receptors, as well as the specific mechanisms they use, remain poorly understood in the brain. This is, in part, because the extraordinary cellular heterogeneity of the brain complicates analysis of the small subpopulations of neurons that mediate sex-specific behaviors.

Having recently identified sex differences in both gene expression and chromatin in brain regions known to regulate sex-specific behaviors, the Tollkuhn lab is now working to understand how hormones generate these molecular sex differences during development through the use of biochemical, genomic, and behavioral analyses. They have developed a method that permits genome-wide analysis of histone modifications or DNA methylation in genetically defined populations of neurons. The team hypothesizes that these epigenetic data, combined with gene expression profiling, define the molecular signature of the critical period for sexual differentiation of the brain. Their goal is to provide a mechanistic link between the transcriptional effects of hormone signaling during development and the consequent social behaviors displayed in adulthood.

Anthony Zador and colleagues study how brain circuitry gives rise to complex behavior. Work in the lab is focused on two main areas. First, they ask how the cortex processes sound, how that processing is modulated by attention, and how it is disrupted in neuropsychiatric disorders such as autism. Recently, the lab found that when a rat makes a decision about a sound, the information needed to make the decision is passed to a particular subset of neurons in the auditory cortex

whose axons project to a structure called the striatum. In the second major line of work in the Zador lab, they are developing new methods for determining the complete wiring instructions of the mouse brain at single-neuron resolution, which they term the "connectome." In contrast to previous methods, which make use of microscopy, these methods exploit high-throughput DNA sequencing. Because the costs of DNA sequencing are plummeting so rapidly, these methods have the potential to yield the complete wiring diagram of an entire brain for just thousands of dollars.

UNDERSTANDING NEURONAL SUBSTRATES OF INTERNAL MODELS AND SENSORIMOTOR TRANSFORMATIONS IN ODOR-DRIVEN BEHAVIORS

F. Albeanu W. Bast D. Cowan D. Hernandez Trejo
H. Chae M. Davis C. Velasquez
Y. Chen P. Gupta C. Wu

The focus of our research group is on understanding how the brain engages in sensorimotor transformations that enable predictions of the sensory outcomes of its actions, such as to make sense of the environment in a closed-loop fashion across different contexts, senses, and brain states.

Toward this end, we use optogenetic methods (such as fast multiphoton laser scanning imaging of genetically encoded neuronal activity reporters or patterned illumination of light-gated neuronal activity switches) coupled with electrophysiological measurements (extracellular and intracellular recordings). We use the rodent olfactory and motor systems as models and monitor neuronal inputs, outputs, and feedback loops across different layers and brain regions.

Understanding Mental Models Underlying Sensorimotor Transformations

During active behaviors, movements shape sensory input and sensory percepts guide subsequent movements. Through experience, the reciprocal relationship between sensory inputs and movements is learned by the brain and further used to predict sensory consequences of upcoming motor actions. Such mental simulations are commonly referred to as internal models. Action-based sensory predictions underlie many important functions, such as cancellation of self-generated percepts or the execution of fast movements in motor control, as well as motor learning. Comparison of sensory predictions to actual observed inputs (i.e., sensorimotor errors) can be used to update the internal model and facilitate motor learning and adaptation. Generation of such sensorimotor predictions requires close coordination between the sensory and motor systems. A widely accepted idea is that feedback signals from motor to sensory areas relay a copy of the motor command, the so-called "efference copy," which is then transformed into sensory coordinates, generating a sensory prediction. However, this idea has received little experimental testing, and most of the mechanistic details remain open questions.

To understand how efferent motor commands are transformed into olfactory sensory predictions, we have developed a closed-loop olfactory localization task for head-fixed mice (*Smellocator*). A head-fixed mouse samples odor delivered from a tube whose opening slides on a horizontal stage in front of the animal (Fig. 1A,B). The mouse can control the placement of the odor source using a lever (Fig. 1C). The animal's task is to center the odor in front of the snout and hold it within the borders of the target zone for 250 msec in return for a water reward. Mice can move the odor source into the correct position within a fraction of a second (Fig. 1D). To ensure that animals do not rely on proprioceptive memory of past trials, we use a set of 12 different transfer functions for each behavioral session (Fig. 1C). Further, in a small fraction of catch trials (10%), we introduced "fake" target zones with water reward at locations far from the snout. In these catch trials, mice opted to place the lever in the correct target zone in front of the snout, and confidently waited there until the end of the trial (~3 sec) even in the absence of water delivery. We probed whether the mice developed a mechanistic understanding of their sensorimotor environment (i.e., an internal model of the world) by applying within-trial perturbations of the odor location. The animals responded by correcting the perturbations within 300–500 msec, successfully bringing the odor source back to the target zone (Fig. 1E). This corrective behavior reveals several interesting points: (1) Mice are continually monitoring the odor location, even after they have

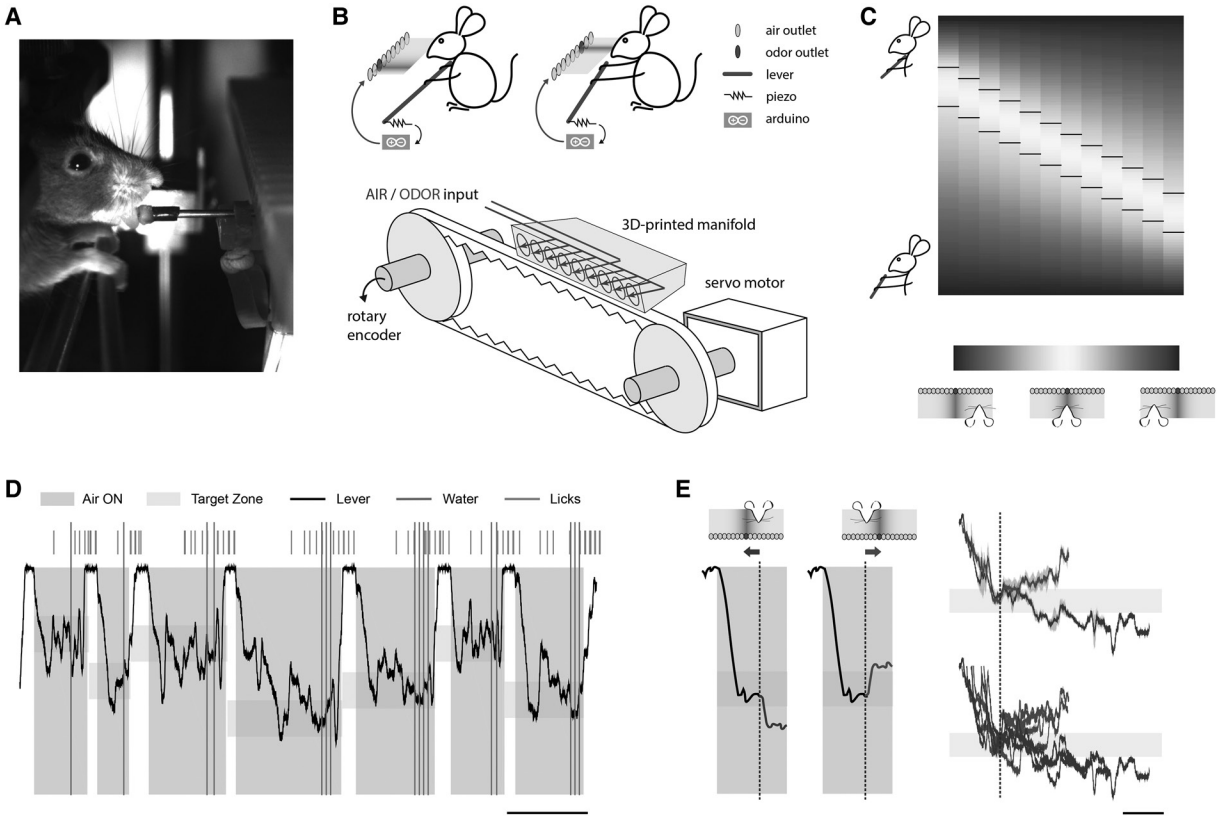


Figure 1. A closed-loop odor localization task (*Smellocator*) to understand the neuronal signature of mental models of sensorimotor transformations. (*A*) Mouse performing the odor locator task. (*B*) Moving the lever engages a servomotor, which controls the location of an odor manifold mounted on a belt in front of the animal's snout. (*C*) Different transfer functions relate the displacement of the lever to the location of the odor source. White: target zone ±2 mm from the snout. (*D*) Behavioral traces from an example session indicating the location of the lever (black), the target zone (horizontal yellow bar), and the licks of the animal (orange ticks as a function of time; scale bar, 1 sec). In different trials (shaded areas), the target zone for the lever is changed randomly such that the odor is centered when the lever is at a different angle. (*E*) Within-trial perturbations of the stimulus lead to a rapid correction by the animal. Blue: leftward, red: rightward shifts in the odor source.

already found the lever location that results in centering the odor; (2) animals can assess whether the odor was displaced leftward or rightward and thus, in principle, obtain a signed estimate of the required sensory correction; and (3) animals can indeed exploit their learned model of the lever action correct for perturbations.

We are currently using high-density electrode probes in both motor areas and across the olfactory processing stream to identify the neuronal signatures and sources of these sensorimotor predictions. We are characterizing the neuronal circuits underlying the computation of sensorimotor errors and learning (i.e., updating of the internal model).

High-Throughput Mapping of Olfactory Bulb Projections across the Brain

Progress in understanding the logic of olfactory coding has been hindered by the inability to map the information flow from glomeruli through downstream brain circuits. Current methods, based on sparse fluorescent labeling, do not allow large-scale characterization of individual neuron projection statistics, nor mapping their functional convergence to distinct target brain areas. The olfactory bulb (OB) relays sensory information represented by olfactory sensory neurons through its output neurons, the mitral and tufted cells (MCs and TCs) to higher brain areas.

The major OB target areas, the piriform cortex (PC), anterior olfactory nucleus (AON), olfactory tubercle, cortical amygdala, lateral entorhinal cortex, and hippocampus, have been proposed to perform distinct computations ranging from odor detection and localization, guiding spatial navigation to odor identification and innate, or learned, stimulus value assignment. It remains largely unknown whether these functional differences emerge locally, or are (in part) dictated by differential inputs from the OB. Using a novel high-throughput sequencing technology (Multiplexed Analysis of Projections by Sequencing, MAPseq), we are investigating the projection patterns of individual MCs and TCs across the brain. In particular, we label a large set of individual MCs and TCs (approximately 500 per experiment) with unique barcoded RNAs using a viral strategy and trace their axonal projections to different target brain areas by sequencing the barcoded RNA. Our approach enables systematic labeling of MCs and TCs across multiple aspects of the bulb that sample inputs from different sets of glomeruli. We cut 200-µm-thick coronal sections along the A-P-axis of fresh frozen brains and capture regions of interest (ROIs) pertaining to the major target areas via laser microdissection. By examining the projection patterns of thousands of MCs and TCs, we investigate whether MCs and TCs broadcast information indiscriminately or have preferential projections to specific brain areas (Fig. 2).

Understanding the Relationship between Olfactory Perceptual Discriminability and Glomerular Response Features

For rodents, the ability to recognize and discriminate particular combinations of volatile compounds is essential for survival. Mice can easily report the difference between weak, similar odors in rich sensory scenes, even when stronger odorants fluctuate in the background. To date, the neural mechanisms underlying such behavior remain unknown.

To understand the neural basis of odor discrimination, we measured and manipulated the activity of the inputs nodes of the olfactory system, the glomeruli. By using wide-field optical imaging in conjunction with odor stimulation, we tracked the position of glomeruli and quantified their odor response properties; this allowed us to define different sets of affine and nonaffine glomeruli with a variable number of components. We aim to determine the relationship between the discriminability of olfactory stimuli and the similarity of glomerular odor response profiles. We additionally quantified the discriminability of the stimuli in relation to the degree of overlap between different sets of glomeruli, as well as the physical separation of glomeruli on the bulb surface.

Toward this end, and to assess the specificity of photo-stimulation, we express red-activatable channelrhodopsin1 (ReaChR) in all mature olfactory

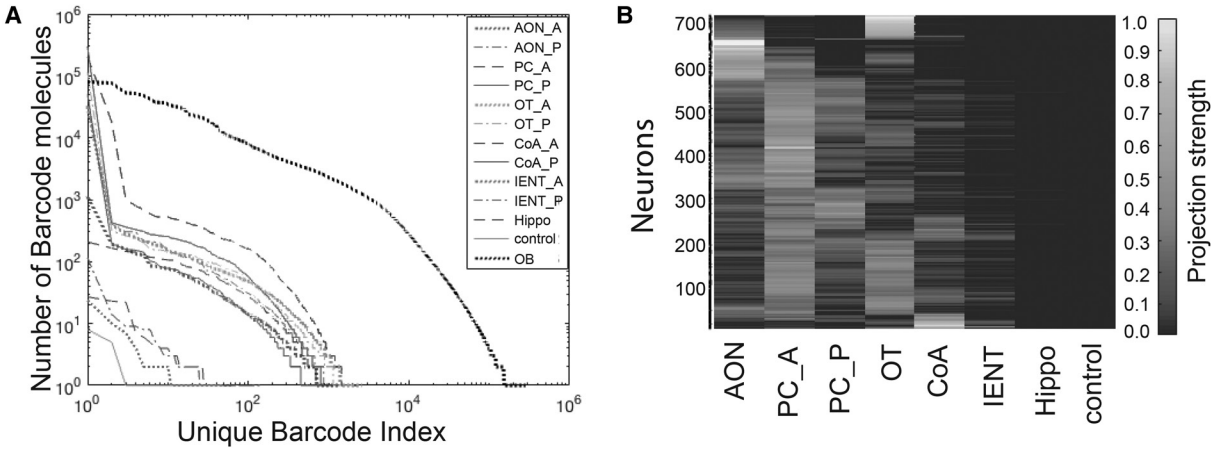


Figure 2. Multiplexed analysis of projections by sequencing (MAPseq) experiments reveal biases in olfactory bulb (OB) projection patterns. (*A*) Distribution of abundance of barcodes in the injection site (OB) and target areas in one example experiment. (*B*) A heat map of all approximately 700 projection patterns from two animals. Abundance of each barcode is normalized to one across the target areas and color-coded.

sensory neurons and GCaMP6f in the OB output neurons. We use digital micro-mirror device (DMD)-based patterned illumination to selectively stimulate combinations of glomeruli on the dorsal surface of the bulb with subglomerular resolution (~10 μm) and high temporal precision (3 msec) in awake, head-fixed mice. Before optogenetic stimulation, using a large odor panel (up to approximately 100 stimuli), we identified the exact locations of glomeruli, revealing their shapes and response tuning to the odors sampled. We further create glomerular light patterns of known odor response similarity (within the range of our panel) and project specific glomerular inputs. In a two-alternative forced-choice discrimination task, we systematically relate the similarity of these light patterns to the perceived difference between them. Further, using a novel strategy to decouple patterned photo-stimulation and two-photon imaging across different axial planes, we are monitoring the responses of MCs and TCs in the deeper layers of the bulb.

We are further implementing strategies that will enable noninvasive, functional dissection of neuronal networks with cellular resolution in behaving animals. This will be brought about via a closed-loop strategy involving real-time control of activity of select neurons with simultaneous monitoring of the concomitant effects of these manipulations on neuronal outputs within the circuit, and elsewhere in the brain. Briefly, we are using digital holography methods via spatial light modulators (SLMs) to optogenetically control neurons of interest at single-cell level and DMD-based methods to control cell type–specific populations across large brain regions (Fig. 3). This allows us to both replicate and systematically manipulate stimulus-evoked activity patterns in a circuit. We are simultaneously using two-photon calcium imaging and electrophysiology within the same and different brain regions (OB vs. olfactory cortex) to dissect how the alteration of select circuit elements, or their specific properties, affects the output of the network. This closed-loop approach will make it possible to determine the spatiotemporal

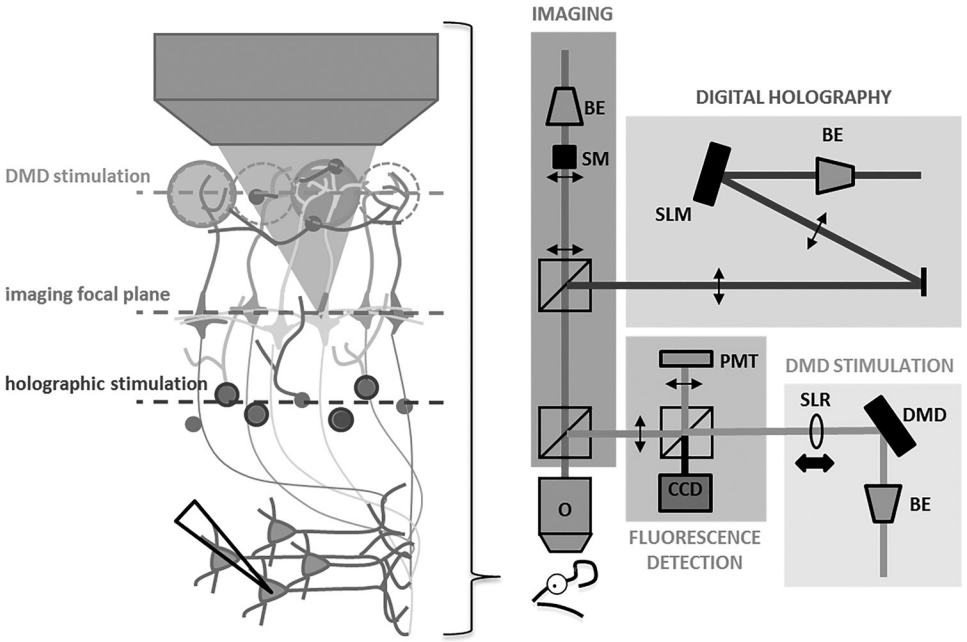


Figure 3. Combined imaging and photo-stimulation setup. (*Left*) Experimental configuration, combining scanning two-photon imaging (red), digital micro-mirror device (DMD) photo-stimulation (blue), and holographic photo-stimulation (brown). (*Right*) Schematic of the experimental microscope. DMD stimulation is used to create spatiotemporal light patterns on the surface (<100 μm). Digital holography is used to photo-stimulate deeper (<500 μm) in the brain with cellular resolution. Calcium activity is monitored in an independent optical plane by two-photon imaging and electrodes in downstream brain regions. BE, Beam expander; SM, scan mirrors; O, objective; PMT, photo-multiplier; SLR, camera lens; DMD, digital micromirror device; SLM, spatial light modulator; CCD, charge-coupled device.

integration rules within the bulb and olfactory cortex, investigate the relevance of spike-time codes, and reveal underlying decoding schemes.

Task-Related Representations in the Cortical-Bulbar Feedback

Sensory systems format information about objects in the environment into neural representations used by the brain depending on context and prior experience. Sensory representations emerge from the interplay between feedforward inputs, as well as local and feedback signals, across brain areas. Given that the output neurons of the main OB—the first processing station of olfactory input—are richly modulated by task contingencies, we asked whether the feedback from PC provides an underlying substrate. To this end, we engage mice in a serial reversal learning task in which a stimulus (odorant or sound) carries distinct motivational value according to different contingencies. In parallel, we monitored the activity of GCaMP5-labeled cortical-bulbar feedback axons. We report the existence of two largely independent populations of axon terminals, responsive in sound versus odor trials, and spanning diverse temporal profiles across both response polarities. Correlation and dimensionality reduction population analysis identified sets of choice-selective and -independent boutons active in both odor and sound trials (Fig. 4). Sound responses occurred mainly during reward (hit) and error (false alarm) trials, further suggesting that the cortical-bulbar feedback carries information on action outcome. Optogenetic inactivation of cortical feedback axons locally in the bulb substantially impaired the task performance compared with controls for both odor and sound trials. These effects were dominated by increased error rates (false alarms) and persisted beyond the light-on trials. In ongoing experiments, we are monitoring the cortical feedback and OB outputs while varying the size of reward, the interval between sensory cues and the time when reward is available, and the identity of odor–sound pairs.

Taken together, our results suggest that cortical feedback conveys information related to task contingencies such as learning rules and reward and is well positioned to flexibly control the bulb output during olfactory behaviors.

Two Parallel Feedforward and Feedback Pathways for Olfactory Information Processing

Areas at the sensory periphery send feedforward signals to the cortex and, in turn, receive massive top-down cortical feedback. To date, the function and degree of specificity of such feedback loops remain poorly understood. In the mammalian olfactory system, the OB output neurons—the MCs and TCs—differ in their intrinsic properties, local connectivity, and projection targets. Two important cortical targets of OB outputs—the AON and the PC—receive feedforward input predominantly from TCs and MCs, respectively. In turn, both AON and PC send massive glutamatergic projections back to local inhibitory OB interneurons (mainly granule cells, GCs). We asked whether and how top-down feedback inhibition received by MCs and TCs is functionally segregated based on their projection targets. We monitored the odor responses of MCs and TCs separately, via multiphoton calcium imaging in awake head-fixed mice, while selectively silencing neuronal activity in the PC or AON. We found that cortical feedback indeed segregated; silencing PC selectively increased the response amplitude and odor correlations of MCs, but not TCs (Fig. 5) (Otazu et al., *Neuron 86*: 1461 [2015]), whereas silencing AON exerted a substantially stronger effect on the odor responses of TCs (increased response amplitude and odor correlations) compared with the MCs. We further investigated whether MCs and TCs process sensory input differentially, and measured their responses across two stimulus dimensions: odorant identity and concentration. MCs showed greater response heterogeneity across odors, as well as concentrations, whereas TCs showed higher population correlations across odorants and nearly monotonic scaling with increasing stimulus concentration. The cross-validated linear (sparse logistic regression) and nonlinear (SVM, polynomial kernel) decoding performance for concentration-invariant odor identity was found to be significantly better from the TC population than from the MC population. Additionally, TC ensembles were superior in generalizing decoding of odor identity when generalizing across concentrations, and could largely solve this simple sensory problem in a feedforward manner. We propose that sensory information in these two parallel loops is formatted differently. Within

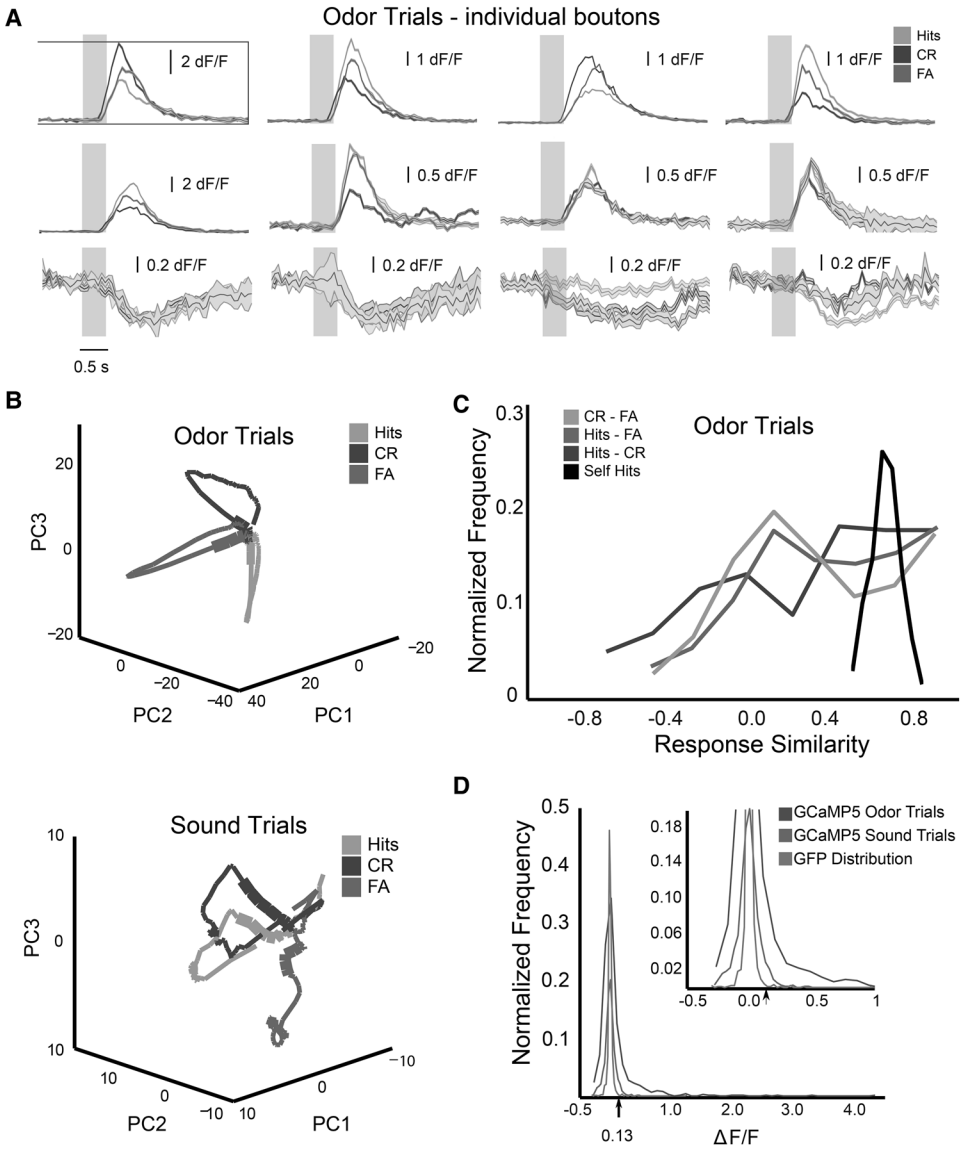


Figure 4. Cortical feedback responses to odor stimuli cluster differently across different task conditions. (*A*) Excitatory and inhibitory trial-averaged responses from individual feedback boutons exemplifying differential modulation across task conditions. Thick line, average response; shaded area ± SEM; gray bar, stimulus (350 msec). (*B*) Ensemble responses of boutons across trial outcomes projected onto the first three principal components (64% of variance). Each trajectory represents the bouton population activity over time for odor (*top*) and sound trials (*bottom*). Origin marks start of trial, and thick line marks the stimulus period. (*C*) Histograms of pairwise response similarity for odor trials across conditions. Colored lines, uncentered correlation between responses of single boutons for different trial outcomes; black line, distribution of self-correlations between two halves of hit trials resampled by bootstrapping 1000 times. (*D*) Histograms of average Δ*F*/*F* for GCaMP responses versus green fluorescent protein (GFP). Each point is the average fluorescence of a 600-msec window starting at the stimulus onset. The arrow denotes the 99th percentile of the GFP distribution, used as the significance threshold for GCaMP responses for further analysis.

this scenario, we are currently testing the hypothesis that cortical targets of TCs (AON, olfactory tubercle) are poised to infer concentration-invariant odor identity, whereas the cortical targets of MCs, such as the PC, are more suited to solving odor-related tasks of a different nature (i.e., contextual learning, flexible rule assignment, generating internal representations of the environment, etc.). In summary, we describe

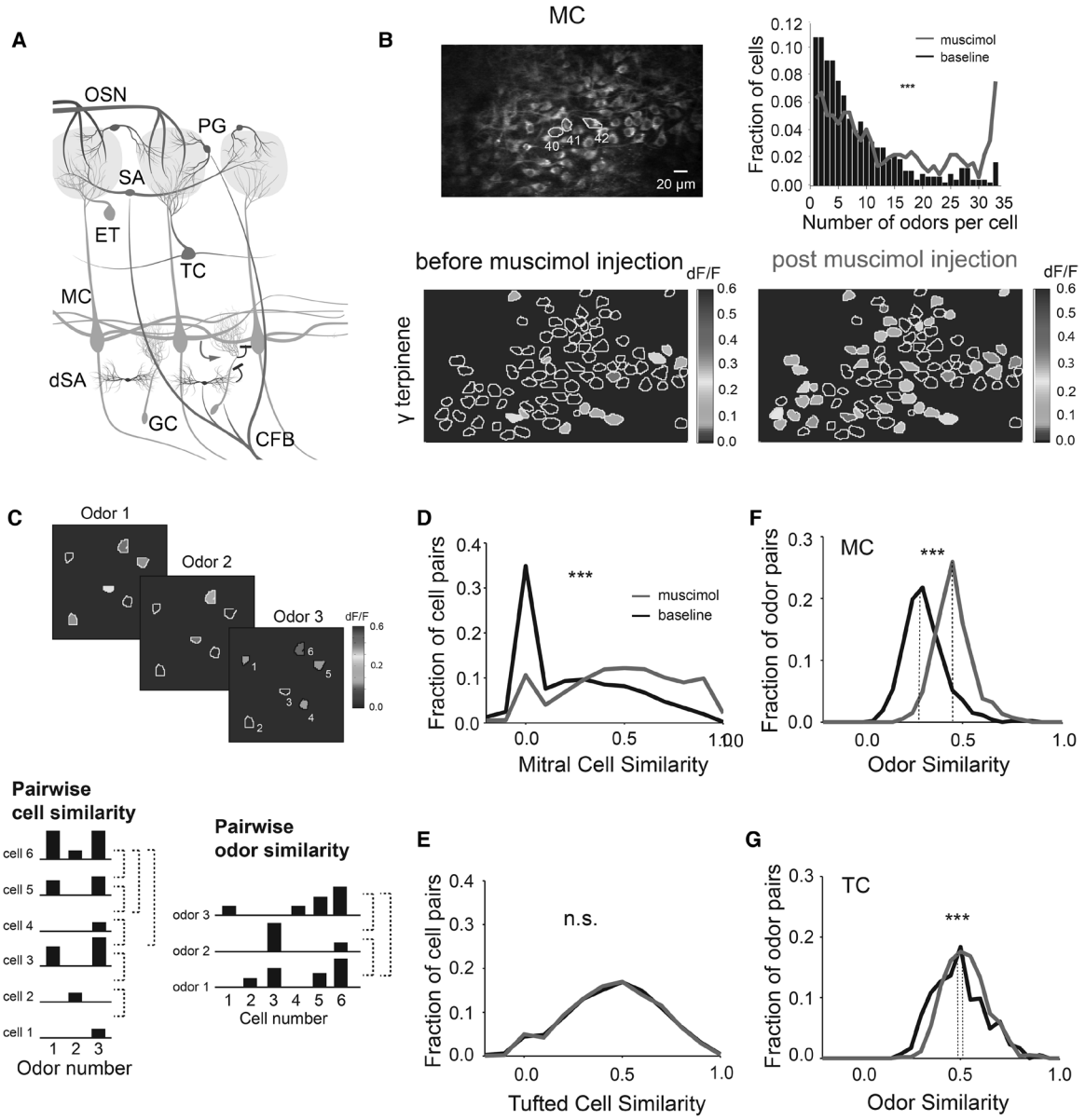


Figure 5. Cortical-bulbar feedback decorrelates the olfactory bulb (OB) output. (*A*) OB circuit and neuronal types; OSN, olfactory sensory neurons; PG, periglomerular cells; SA, superficial short axon cells; ET, external tufted cells; TC, tufted cells; MC, mitral cells; dSA, deep short axon cells; GC, granule cells; CFB, cortical feedback fibers synapsing onto inhibitory interneurons across the OB. (*B*) (*Top left*) Average resting fluorescence of an example field of view in the mitral cell (MC) layer (~220 µm from surface). (*Bottom*) Color map showing average fluorescence change in response to γ terpinene in the field of view before (*left*) and after (*right*) muscimol injection (injected in anterior piriform cortex [PC] to suppress cortical activity and, thus, disable cortical feedback). (*Top right*) Histogram of the number of odors individual MCs responded to before (black bars) and after (red trace) muscimol injection; *** indicates significance level ($p < 0.001$, Wilcoxon signed rank test). (*C*) Schematic exemplifying pairwise cell and pairwise odor similarity calculations for a given field of view. (*Left*) Cartoon showing responses of six identified regions of interest (ROIs, yellow outlines) within a given field of view across three odors; color indicates the average response amplitude (*dF/F*) for each ROI. (*Center*) An odor response spectrum (ORS) is calculated for each ROI (cell) as the vector containing the average *dF/F* for each odor; pairwise cell similarity is calculated as the uncentered correlation between the ORS vectors for each pair of cells (indicated by dotted lines). (*Right*) A cell response spectrum (CRS) is calculated for each odor as the vector containing the average *dF/F* for each cell on presentation of the given odor; pairwise odor similarity is calculated as the uncentered correlation between the CRS vectors for each pair of odors. (*D–G*) Histogram of pairwise cell (*left*) and odor (*right*) similarities of mitral cells (MCs) and tufted cells (TCs) before (black, baseline) and after (red) muscimol injection; *** indicate significance level ($p < 0.001$, Wilcoxon signed rank test); dotted lines indicate the median.

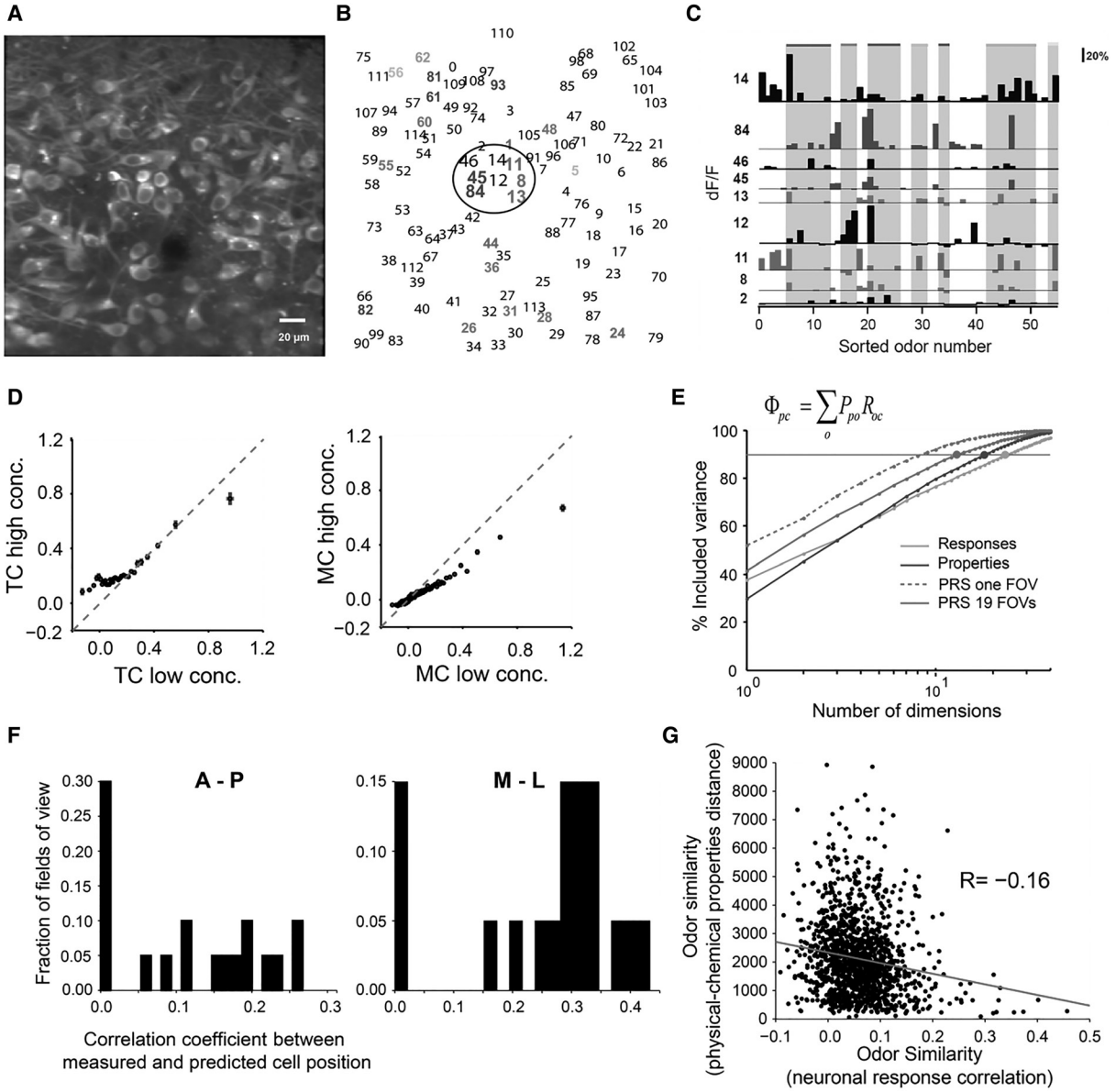


Figure 6. Mapping odor space on mitral cell (MC) activity patterns: (*A*) Average resting fluorescence multiphoton image in the MC layer. (*B*) Numbers indicate the relative positions of 112 MC bodies in the imaged field of view. (*C*) Example odor response spectra of MC bodies within an arbitrarily picked 75-µm-diameter region in *A*, sorted according to functional chemical groups (tiglates, thiazoles, ethyl esters, aldehydes, ketones, alcohols, and acids). (*D*) Scatter plots showing odor-induced change in TCs (*left*) and MCs (*right*) with increasing odor concentration. (*E*) Dimensionality of odorant properties, MC and TC neuronal responses and MC and TC receptive fields (properties response strength, PRS, Φ). (*F*) A LASSO/jackknife predictor yields a set of AP and ML cell positions that are weakly correlated with observed values. (*G*) Odor pairwise similarity across a set of 1660 physical–chemical properties versus neuronal responses.

two parallel feedforward and -back streams of odor processing in the mammalian brain with differential inhibitory organization and neural representation for odor identity and concentration, further providing a testable framework during olfactory behaviors that go beyond sensory decoding.

Mapping Odor Space onto Neuronal Representations in the Olfactory System

Unlike many other sensory systems, low-dimensional metrics for characterizing stimuli have remained elusive for olfaction, and it is unclear what features of chemical stimuli are represented by neurons. We would like to relate neural activity in the early olfactory system of mice to the physical–chemical properties of odorants. The elementary stimulus features encoded by the olfactory system remain poorly understood. We examined the relationship between 1666 physical–chemical descriptors of odors and the activity of OB inputs, as well as outputs in awake mice. Glomerular cell and MC and TC responses were sparse and locally heterogeneous, with only a coarse dependence of glomerular positions on physical–chemical properties. Odor features represented by ensembles of MCs and TCs were overlapping, but distinct from those represented in glomeruli, consistent with extensive interplay between feedforward and -back inputs to the bulb. This reformatting was well described as a rotation in odor space. The descriptors accounted for a small fraction in response variance, and the similarity of odors in physical-chemical space was a poor predictor of similarity in neuronal representations. Our results suggest that commonly used physical–chemical properties are not systematically represented in bulbar activity and encourage further search for better descriptors of odor space (Fig. 6) (collaboration with the Koulakov lab).

Other Collaborative Projects with CSHL Groups

Huang: DLP-based patterned stimulation to functionally map the motor cortex.

Koulakov, Lee, Zador: Sequencing the OB—bridging the gap between glomerular odor responses and odor receptor sequences by identifying the molecular identity of glomeruli.

Li: Fiber-optic-based approach to monitor neuronal activity in punishment and reward neuronal circuits during behavior. We are focusing on monitoring and manipulating activity in the lateral division of central amygdala and the insula cortex.

Osten: Developing light-sheet-based approaches for fast optical reconstruction of neuronal circuits. *Zador:* Optical monitoring and manipulation of neuronal activity in genetically and anatomically defined cortical circuits in animal models of cognition.

INVESTIGATING NEURAL CIRCUITS FOR MULTISENSORY DECISION-MAKING

A. Churchland G. Bekheet, L. Chartarifsky, S. Gluf, M. Kaufman, E. Lu, S. Musall, F. Najafi, O. Odoemene, S. Pisupati

Making use of sensory information requires more than simply relaying incoming signals from the sensory organs. It requires interpreting information, classifying it, drawing inferences, and ultimately using the context of behavioral goals to make a decision about its meaning. A decision is a commitment to a particular choice or proposition at the expense of competing alternatives. In some situations, decisions involve integration of evidence—that is, they combine multiple pieces of information from the environment or from memory. These decisions can provide a framework in which to investigate complex cognitive processes and open a window into higher brain function.

Although previous experiments have begun to reveal how neural systems combine evidence to make decisions, they have left a critical gap in our understanding. Specifically, very little is currently known about the neural mechanisms that make it possible to combine information from multiple sensory modalities for decisions. The gap is apparent, although it is clear from behavioral observations that neural systems can combine multisensory information: When parsing speech in a crowded room, for example, the listener makes use of both auditory information (the speaker's vocal sounds) and visual information (the speaker's lip movements). Understanding the neural mechanisms of multisensory decisions is critical for two reasons. First, it is essential for a complete understanding of sensory perception because real-world stimuli rarely activate a single sense in isolation. Therefore, understanding how the brain interprets incoming information requires understanding how the brain merges information from different senses. Second, it is likely of clinical importance: Several developmental abnormalities appear to be related to difficulties in integrating sensory information. For example, abnormalities in multisensory processing are a hallmark of subjects with Autism Spectrum Disorder. Impairments in multisensory processing are also observed in subjects with a collection of sensory abnormalities, known together as sensory processing disorder, and may also be evident in patients with Rett syndrome and dyslexia. Understanding the neural mechanisms of multisensory integration could inform treatment of those conditions. Our long-term goal is to understand how the brain can make decisions that integrate inputs from our multiple senses, stored memories, and innate impulses. Our projects in previous years constituted the very first steps toward achieving this goal: We developed a multisensory behavior that could be precisely quantified in both humans and rodents (Raposo et al., *J Neurosci 32:* 3726 [2012]; Sheppard et al., *J Vision 13:* 4 [2013]). In addition, we built on the foundation by measuring the responses of neurons in rodents engaged in the behavior, affording insight into the neural circuits underlying multisensory decisions (Raposo et al., *Nat Neurosci 17:* 1784 [2014]).

Microcircuits within Posterior Parietal Cortex that Support Decision-Making

L. Chartarifsky, E. Lu, S. Pisupati [in collaboration with M. Kaufman, Stanford University]

There are two projects within this category. The goal of the first project is to gain a deeper understanding of the neural circuits that allow auditory, visual, and multisensory decision-making. This year, we finalized a project that used optogenetics to manipulate neural activity and, thus, understand the causal role of a candidate decision-making structure, the posterior parietal cortex (PPC) (Licata et al. 2017). This paper revealed that the PPC plays a causal role in visual decisions, but not in auditory decisions or in multisensory integration.

To identify the key region that supports multisensory integration, Lital Chartarifsky and Sashank Pisupati are performing brain-wide manipulations of

candidate areas. Their approach is to disrupt neural activity in a candidate area and look at the effects on auditory, visual, and multisensory decisions. They have developed a model-based system to evaluate and interpret the behavioral data. This model-based system is a major step forward because connecting decision-making to its underlying behavior is a notoriously difficult problem (Churchland and Kiani, *Curr Opin Behav Sci 11:* 74 [2016]).

Ethological Decision-Making

G. Bekheet [in collaboration with A. Juavinett, University of California, San Diego]

Here, we aim to evaluate neural activity during a naturally occurring behavior in which mice flee when presented with an overhead "looming" stimulus. We are recording neural activity during this behavior using high-density electrodes called Neuropixel probes, which will allow us to measure electrical signals from approximately 150 neurons simultaneously. We have evaluated the animal's sensitivity to visual stimuli of varying contrasts and compared behavioral responses with visual and multisensory stimuli. The use of ethological stimuli in this project offers many advantages (Juavinett et al. 2017): Animal training is unnecessary, speeding up the experimental pipeline. Further, studying ethological behavior provides a glimpse of neural activity that supports behaviors that evolved over many generations to serve a specific function for the animal.

Population Dynamics of Neurons during Decision-Making

This work was done in collaboration with M. Kaufman (Stanford University) and F. Najafi (University of Pennsylvania).

The goal of this project is to understand how population activity changes from one neural structure to the next to support behavior. To tackle this question, we use two-photon imaging to measure the responses of 300 to 400 neurons simultaneously. This approach will revolutionize the kinds of questions about decision-making that we can address, especially when used in conjunction with emerging mathematical techniques for analysis (Churchland and Abbott, *Nat Neurosci 19:* 348 [2016]).

Farzaneh Najafi is using this technique to study the dynamics of populations of neurons during decision-making in mice. She is focusing on how excitatory and inhibitory neurons together contribute to evolving population responses. Her approach will include a consideration of the animal's strategy, such as the degree to which it uses the outcome of previous trials to guide current decisions.

Both projects have benefitted from technical support from the Albeanu lab, our neighbors in the Marks building, and valued collaborators both at Cold Spring Harbor and at Columbia University.

Brain-Wide Macrocircuits that Support Decision-Making

S. Gluf, O. Odoemene, S. Pisupati [in collaboration with M. Kaufman, Stanford; S. Musall, Universität Zurich]

The goal of this project is to understand how multiple neural areas make up brain-wide macrocircuits that support decision-making. The project has two key components.

The first component was led by Onyekachi Odoemene, a Watson School student who graduated in September 2017. He leveraged inactivation methods to determine the role of multiple visual areas in transforming sensory information into signals that guide action (Odoemene et al. 2017).

The second component is led by Simon Musall. Simon uses a combination of wide-field imaging and two-photon imaging to understand which neural areas are active during different moments of a perceptual decision. To do so, he has been developing new behaviors in mice. Research technician Steven Gluf has provided support for this project.

International Brain Laboratory

Dr. Churchland cofounded this organization and secured funding in 2017 from the Wellcome Trust and the Simons Collaboration on the Global Brain. The International Brain Laboratory is a virtual laboratory unifying a group of 21 experimental and theoretical neuroscience groups distributed across the world to understand the neural computations supporting decision-making (International Brain Laboratory, *Neuron 96:* 1213 [2017]).

PUBLICATIONS

Juavinett AL, Erlich JC, Churchland AK. 2017. Decision-making behaviors: Weighing ethology, complexity, and sensorimotor compatibility. *Curr Opin Neurobiol* **49:** 42–50.

Licata AM, Kaufman MT, Raposo D, Ryan MB, Sheppard JP, Churchland AK. 2017. Posterior parietal cortex guides visual decisions in rats. *J Neurosci* **37:** 4954–4966.

Odoemene O, Pisupati S, Nguyen H, Churchland AK. 2017. Visual evidence accumulation guides decision-making in unrestrained mice. bioRxiv doi: 10.1101/195792.

STRUCTURAL NEUROPHARMACOLOGY OF ION CHANNELS

H. Furukawa E. Chou M. Regan N. Simorowski J. Syrjanen N. Tajima J.X. Wang

The research in the Furukawa lab attempts to understand molecular events that initiate cellular signal transductions involved in neurotransmission and neuroplasticity in the mammalian brain with a scope to develop therapeutic compounds for treatment of neurological diseases and disorders, including schizophrenia, depression, stroke, and Alzheimer's disease. To achieve our goals, we conduct structural and functional studies on ion channels that control intracellular calcium signaling on stimulation by voltage and/or neurotransmitters. Those ion channels regulate strength of neurotransmission, the fundamental process for neuronal communication. Dysfunction of the ion channels studied in our group is highly implicated in neurological disorders and diseases noted above. The abnormal activation of the ion channels is caused by a number of factors, including excessive transmission of neurotransmitters and point mutations in the ion channels, that alter their functional properties. To understand functions of normal and abnormal ion channels, we use structural biology techniques, including X-ray crystallography and single-particle cryo electron microscopy (cryo-EM), to determine three-dimensional atomic structures of target ion channels and test structure-based mechanistic hypotheses by site-directed mutagenesis in combination with biochemical and biophysical techniques such as electrophysiology. In 2017, we advanced our understanding of how *N*-methyl-D-aspartate receptor (NMDAR) recognizes natural and synthetic ligands. NMDA receptors form obligatory heterotetrameric ion channels that are composed of GluN1 subunits (eight splice variants: 1–4a or b) and GluN2 subunits (A–D), and/or GluN3 subunits (A or B).

Zinc Inhibits NMDA Receptor Activity

Zinc has emerged as a major modulator of brain functions in mammals in recent years. It is mostly present at axonal termini of various sets of glutamatergic neurons and is cotransmitted with L-glutamate in an activity-dependent manner (Westbrook and Mayer 1987; Vogt et al. 2000; Vergnano et al. 2014). The released zinc modulates activity of a number of ion channels, including NMDA receptors. The extracellular zinc binds and inhibits the NMDA receptors that contain GluN2A and GluN2B in a voltage-independent and allosteric manner at nanomolar and micromolar potencies, respectively (Chen et al. 1997; Paoletti et al. 1997; Low et al. 2000; Rachline et al. 2005). The inhibition with the nanomolar potency or the high-affinity zinc inhibition in the GluN2A-containing NMDA receptors play a critical role in controlling pain sensation (Nozaki et al. 2011). The structural studies of NMDA receptors have advanced moderately in recent years (Regan et al. 2015). The crystal structures of the isolated GluN2B ATD and the heterodimer of GluN1b and GluN2B ATDs (GluN1b-GluN2B ATD) showed that the micromolar zinc-binding site and the ifenprodil-binding site reside at the cleft of the bilobed architecture of the GluN2B ATD (Karakas et al. 2009) and the subunit interface of the GluN1b-GluN2B ATD (Karakas et al. 2011), respectively.

The recent studies on the intact NMDA receptors showed conformational movement of the GluN1-GluN2B NMDA receptors in various functional states stabilized by different combinations of ligands (Karakas and Furukawa 2014; Lee et al. 2014; Tajima et al. 2016; Zhu et al. 2016). In contrast to the GluN1-GluN2B subtype, structural studies on the GluN1-GluN2A subtype, except on LBD (Furukawa et al. 2005; Hansen et al. 2013; Jespersen et al. 2014; Hackos et al. 2016; Volgraf et al. 2016), have lagged because of technical difficulties associated with protein production. Consequently, the field has lacked any means to structurally compare NMDA receptor subtypes and pinpoint subtype-specific elements in ATD. Thus, we solved the crystal structure of the heterodimers of GluN1b and GluN2A ATDs (GluN1b-GluN2A ATD) in the presence and absence of zinc.

Bilobed Architectures of GluN1b and GluN2A ATDs

We developed a few novel methods to recombinantly express and purify the GluN1b-GluN2A ATD proteins. Crystallization required a Fab fragment of the GluN1-targeting antibody. The crystal structures of the GluN1b-GluN2A ATD show an ATD heterodimer and a Fab in the asymmetric unit, which reinforces the view that the heterotetrameric subunit arrangement of the intact GluN1-GluN2A NMDA receptor is likely a dimer of two GluN1-GluN2A heterodimers as previously shown for the GluN1-GluN2B NMDA receptors (Karakas and Furukawa 2014; Lee et al. 2014). Both the GluN1b ATD and the GluN2A ATD have bilobed architectures composed of upper (R1) and lower (R2) domains (Fig. 1). Zinc binds to the R1–R2 interface of the GluN2A ATD and stabilizes the bilobe in a "closed" conformation. Although there are minor differences locally, the main chain carbons (Cαs) of the three GluN1b-GluN2A ATD structures (Zn1, Zn2, and EDTA) are superimposable to one another with root-mean-square deviation (RMSD) of 0.802 Å, 0.539 Å, and 0.962 Å, between Zn1 and Zn2, Zn1 and EDTA, and Zn2 and EDTA, respectively. The structure of the GluN1b ATD in the GluN1b-GluN2A ATD is almost identical to that observed in the crystal structures of the GluN1b-GluN2B ATD bound to ifenprodil (ifenprodil-GluN1b-GluN2B ATD) (Karakas et al. 2011) with RMSD of 0.696 Å over 347 superimposable Cαs. In contrast, the GluN2A ATD cannot be superimposed onto the zinc-bound GluN2B ATD (Zn-GluN2B ATD) (Karakas et al. 2009) while the overall bilobed architectural feature is conserved, whereas their R1 and R2 regions can be superimposed individually with RMSD values of 1.54 Å and 1.12 Å, respectively. This stems from the major difference in the extent of the R1–R2 separation between the Zn1-GluN2A ATD and the Zn-GluN2B ATD in which the bilobed architecture of the Zn1-GluN2A ATD is ~13° more open compared with that in the Zn-GluN2B ATD (Karakas et al. 2009). That is, the bilobe of the GluN2A ATD is not capable of closing as much as the GluN2B ATD.

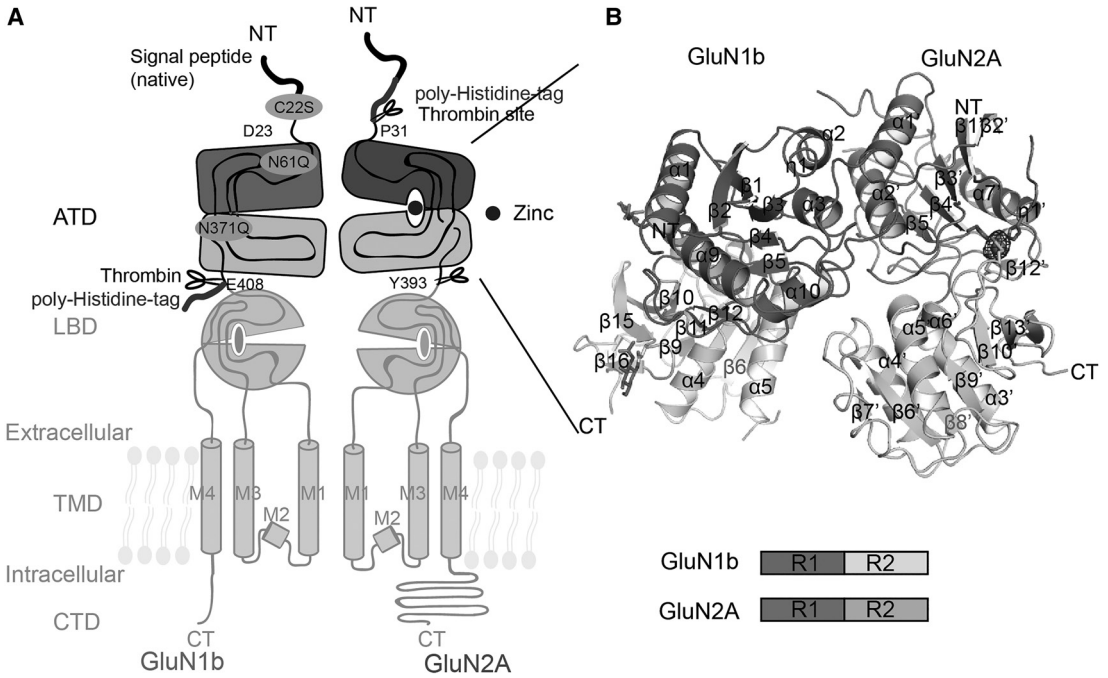


Figure 1. Crystal structure of GluN1b-GluN2A ATD. (*A*) Domain organization of the GluN1b-GluN2A NMDA receptors. (*B*) The crystal structure of GluN1b-GluN2A ATD heterodimer complexed to zinc showing bilobed architecture composed of R1 and R2 domains. The Fab fragment used to assist crystallization and contained in the asymmetric unit is omitted for clarity.

Lack of Phenylethanolamine-Binding Pocket in the GluN1b–GluN2A ATD Interface

The heterodimeric GluN1–GluN2B subunit interface within ATD is where phenylethanolamine compounds, such as ifenprodil, bind in the GluN1-GluN2B subtype-specific manner (Karakas et al. 2011); however, the mechanism of this subtype-specific binding has remained obscure because of lack of ATD structures from any NMDA receptor subtypes except the GluN1-GluN2B. Although absence of binding of ifenprodil to the GluN1-GluN2C and GluN1-GluN2D subtypes may be explained by nonconservation of primary sequences, all but one residue, GluN2A-Met112, is conserved in GluN2A at the region corresponding to the ifenprodil binding site (Karakas et al. 2011). We have previously altered GluN2A-Met112 to the corresponding residue in GluN2B, isoleucine, and observed no gain of ifenprodil sensitivity (Karakas et al. 2011). Here, our structure shows that the ~12° intersubunit rotation in the GluN1b-GluN2A ATD compared with the ifenprodil-GluN1b-GluN2B ATD described above results in shortening the distance between GluN1-α3 and GluN2-α2′, thereby eliminating the cavity space for compound binding (Fig. 2). Indeed, the GluN1b-GluN2A ATD structures show

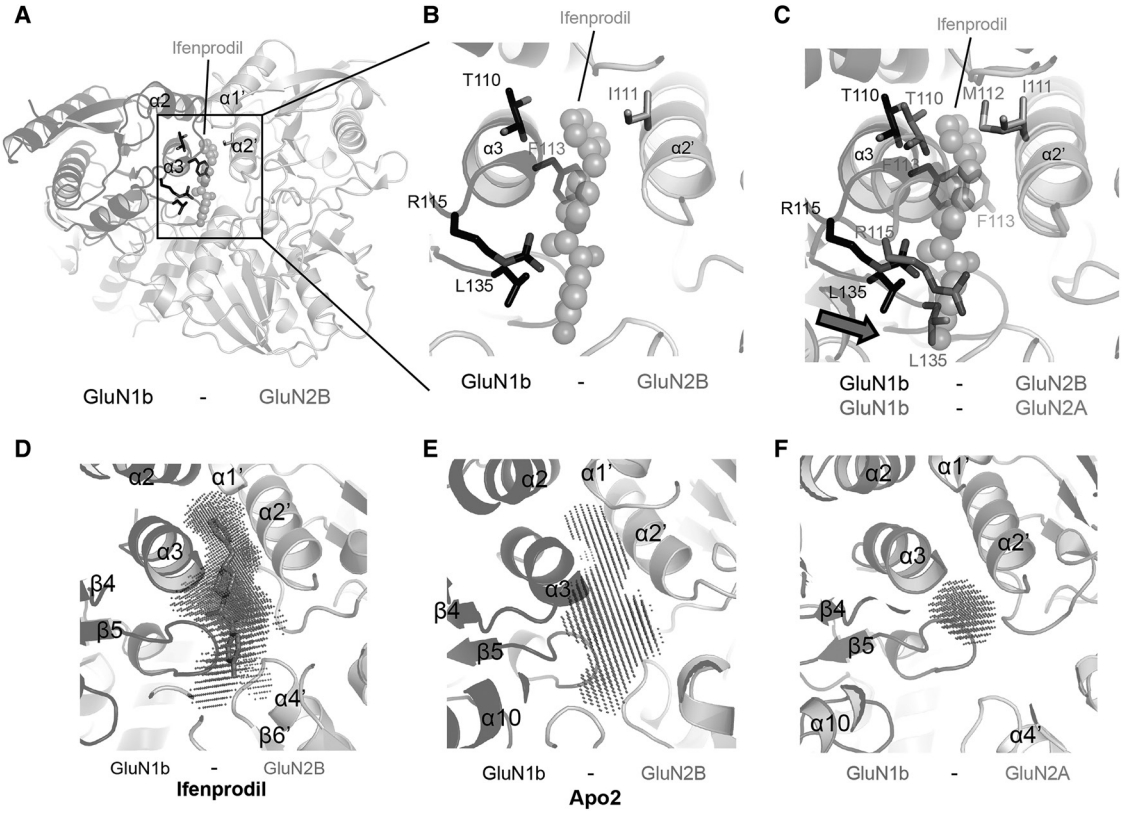


Figure 2. Comparison of the GluN1–GluN2A and GluN1–GluN2B subunit interfaces. (*A,B*) Ifenprodil-binding pocket at the GluN1b–GluN2B subunit interface (PDB code: 3QEL). (*C*) Superposition of the Zn-GluN1b-GluN2A ATD and the ifenprodil-GluN1b-GluN2B ATD at the R1 lobes of GluN2A and GluN2B show that the ifenprodil-binding pocket is "filled in" by the residues from GluN1b in the GluN1b-GluN2A ATD heterodimer because of the smaller gap between the two subunits (arrow). Volume of the protein cavities at the subunit interface is calculated for the ifenprodil-GluN1b-GluN2B ATD (PDB code: 3QEL) (*D*), apo2-GluN1b-GluN2B ATD (*E*), and the Zn1-GluN1b-GluN2B ATD (*F*) by the software KVFinder (Oliveira et al. 2014). Protein cavities are represented as dots. The interface cavity in the GluN1-GluN2B ATD is sufficiently large to accommodate ifenprodil (*D,E*; volume 697 and 550 Å^3), whereas that in the GluN1b-GluN2A ATD is not (*F*; volume ~ 130 Å^3).

a cavity with insufficient volume (~130 Å^3) (Fig. 2) to accommodate the ifenprodil molecule, which has a volume size of 324 Å^3. The measured volume of the intersubunit cavity is 697 Å^3 (Fig. 2D) in the ifenprodil-GluN1b-GluN2B ATD.

To determine whether the cavity in the subunit interface in the GluN1b-GluN2B ATD is inherently present or formed exclusively by binding of ifenprodil, we obtained a crystal structure of the GluN1b-GluN2B ATD in the absence of ifenprodil. This structure (apo2-GluN1b-GluN2B ATD) is distinct from our recently published apo-GluN1b-GluN2B ATD (Tajima et al. 2016) in that the bilobe of GluN2B ATD is closed and the inter-GluN1-GluN2 subunits are arranged in a manner similar to the ifenprodil-GluN1b-GluN2B ATD, which is an indication that the conformation of the GluN1b-GluN2B ATD can intrinsically fluctuate as previously predicted (Gielen et al. 2009). Importantly, this new apo-state structure with the closed bilobe (apo2-GluN1b-GluN2B ATD) retains the subunit interface cavity with sufficiently large volume (550 Å^3) to accommodate ifenprodil, showing that the presence of the large cavity is intrinsic to the GluN1b-GluN2B ATD (Fig. 2E). In contrast, the apo-GluN1b-GluN2B ATD with the "open" GluN2B ATD bilobe (Tajima et al. 2016) has little or no cavity space, indicating that the ifenprodil-binding site is created only when the GluN2 ATD bilobe is sufficiently closed.

High-Affinity Zinc-Binding Site within GluN2A ATD

Our structural analysis unambiguously reveals the molecular organization of the zinc-binding site responsible for the high-affinity zinc inhibition. Zinc anomalous difference Fourier maps unambiguously show one zinc-binding site at the inter-R1-R2 cleft of the GluN2A ATD in both the Zn1-GluN1b-GluN2A ATD and the Zn2-GluN1b-GluN2A ATD. The electron density for the zinc-binding site within GluN2A ATD unambiguously shows that zinc is coordinated by the four residues, GluN2A-His44, -His128, -Glu266, and -Asp282 (Fig. 3A,B). The zinc-recognition pattern in the GluN2A ATD is different

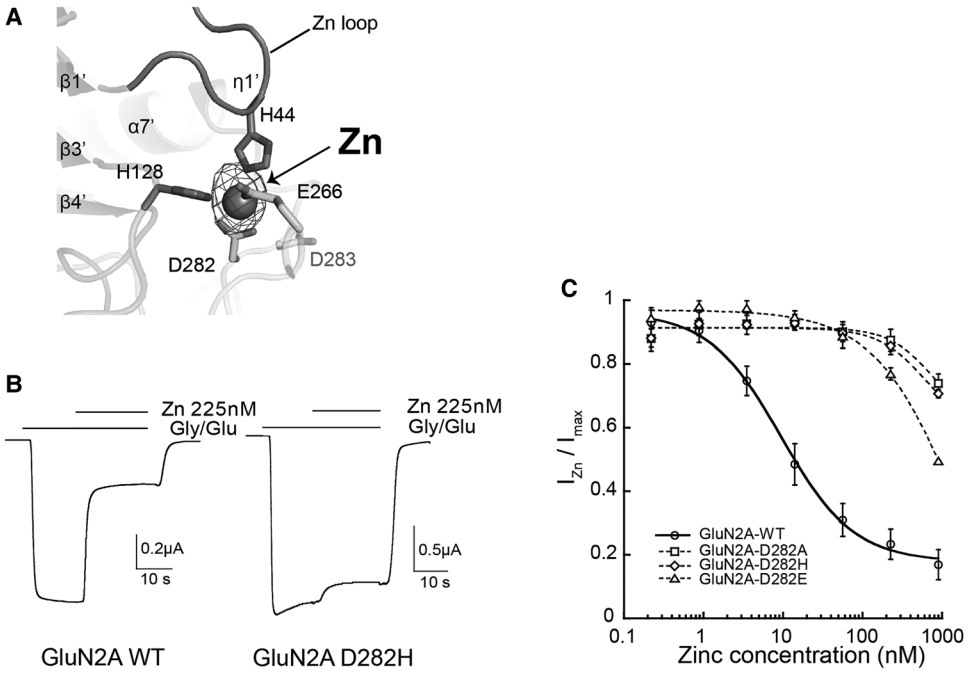


Figure 3. The high-affinity zinc-binding site of GluN2A. (*A*) The high-affinity zinc-binding site of GluN2A where zinc (gray sphere) is coordinated by GluN2A-His44, -His128, -Glu266, and -Asp282. The mesh represents the zinc anomalous difference Fourier map contoured at 5.0σ. (*B*) The high-affinity zinc inhibition is abolished by mutation of GluN2A-Asp282. Shown here are representative recordings of zinc inhibition of the WT GluN2A and GluN2A-Asp282His coexpressed with the WT GluN1-1a by TEVC held at –60 mV and the data plot (*C*).

from the one previously observed in the GluN2B ATD with only two directly coordinating residues (Fig. 3D) (Karakas et al. 2009). Notably, the loop that contains GluN2A-His44 (Zn-loop) has a length and structure distinct from the GluN2B ATD (Fig. 3C–E). Extensive mutagenesis studies in the past showed that altering GluN2A-His44, GluN2A-His128, and GluN2A-Glu266, but not GluN2A-Asp282, to alanine reduces potency and efficacy of the zinc inhibition (Choi and Lipton 1999; Fayyazuddin et al. 2000). Instead, the GluN2A-Asp283Ala mutation was shown to alleviate the high-affinity zinc inhibition, which led to the conclusion that GluN2A-Asp283 but not GluN2A-Asp282 is involved in direct zinc coordination (Stroebel et al. 2011). To validate our structural observation that GluN2A-Asp282 is directly involved in coordination of zinc and therefore high-affinity zinc inhibition, we mutated GluN2A-Asp282 to alanine, histidine, and glutamate and measured zinc inhibition by electrophysiology and found that site-directed mutation of GluN2A-Asp282 dramatically affected zinc inhibition in our hands (Fig. 3F,G). Together, the combination of our mutagenesis results and the previous results (Choi and Lipton 1999; Fayyazuddin et al. 2000; Low et al. 2000; Paoletti et al. 2000) support the structural observation that side chains of GluN2A-His44, GluN2A-His128, GluN2A-Glu266, and GluN2A-Asp282 are direct coordinators of zinc.

To understand a plausible pattern of zinc association and dissociation with the GluN2A ATD, zinc in the Zn1-GluN1b-GluN2A ATD was removed by exhaustively soaking crystals against the crystallization solution containing 1 mM EDTA (EDTA-GluN1b-GluN2A ATD). As a quality control of the experiment, we collected X-ray diffraction data at the zinc peak wavelength (1.28 Å) and observed no signal in the anomalous difference Fourier map to confirm absence of zinc in the crystal. Although the crystal packing kept the overall bilobe conformation to that observed in the Zn1- or Zn2-GluN1b-GluN2A ATD, the EDTA-GluN1b-GluN2A ATD revealed a significant structural change in the Zn-loop that contains the zinc coordinating residue, GluN2A-His44, uniquely present in GluN2A (Fig. 3). The Zn-loop is ordered and directed toward the zinc-binding site at the cleft in the Zn1- or Zn2-GluN1b-GluN2A ATD. In the EDTA-GluN1b-GluN2A ATD, the Zn-loop flaps away from the bilobe cleft and the electron density for the side chain of GluN2A-His44 becomes disordered. This result implies that the high-affinity zinc binding follows an induced-fit mechanism involving rearrangement of the Zn-loop. It is noteworthy that this local change in the Zn-loop conformation is not caused by the crystal packing. Despite extensive efforts, we were unable to capture the bona fide apo-state with a plausible open-cleft ATD as we recently observed in the GluN2B ATD (Tajima et al. 2016) because of an unsuccessful attempt to crystallize the GluN1b-GluN2A ATD proteins in the complete absence of zinc.

PUBLICATIONS

Amin JB, Salussolia CL, Chan K, Regan MC, Dai J, Zhou HX, Furukawa H, Bowen ME, Wollmuth LP. 2017. Divergent roles of a peripheral transmembrane segment in AMPA and NMDA receptors. *J Gen Physiol* **149:** 661–680.

Elegheert J, Cvetkovska V, Clayton AJ, Heroven C, Vennekens KM, Smukowski SN, Regan MC, Jia W, Smith AC, Furukawa H, et al. 2017. Structural mechanism for modulation of synaptic neuroligin-neurexin signaling by MDGA proteins. *Neuron* **96:** 242–244.

Romero-Hernandez A, Furukawa H. 2017. Novel mode of antagonist binding in NMDA receptors revealed by the crystal structure of the GluN1-GluN2A ligand-binding domain complexed to NVP-AAM077. *Mol Pharmacol* **92:** 22–29.

In Press

Regan MC, Grant T, McDaniel MJ, Karakas E, Zhang J, Traynelis SF, Grigorieff N, Furukawa H. 2018. Structural mechanism of functional modulation by gene splicing in NMDA receptors. *Neuron* **98:** 521–529.

ASSEMBLY AND ORGANIZATION OF CORTICAL CIRCUITS CONTROLLING VOLITIONAL MOVEMENTS

J. Huang X. An, S.W. Galbavy, D. Huilgol, S. Kelly, G. Kim, J. Levine, J. Lu, K. Matho, H. Mohan, A. Paul, Y. Qian, R. Raudales, P. Wu

The neocortex consists of a constellation of functional areas that form a representation map of the external and internal worlds. These areas constitute an interacting network that integrates multisensory information with emotional drive and internal goals to coordinate behavior. The fundamental organization plan of the cortex is conserved across mammalian species and is ultimately encoded in the genome, which directs developmental genetic programs to assemble species-stereotyped features, such as the basic layout of areas and scaffold of areal networks, laminar architecture, and local circuit templates. Central to the developmental assembly and functional organization of cortical circuits is the generation of a large set of cardinal neuron types: the glutamatergic pyramidal neurons (PyNs), which form myriad cortical processing streams and output channels, and the GABAergic interneurons, which assemble local connectivity motifs. The broad and overarching theme of my research program is to understand the principles underlying the assembly and organization of cortical circuits by integrating multifaceted studies of cortical neuron types in the functional context of cortical control of volitional movement in the mouse. To achieve this, we combine genomic, genetic engineering, anatomical, physiological, and behavioral approaches. Although highly ambitious, such a multifaceted research program is, in fact, necessary to truly understand the cellular basis of brain circuit organization. With rapid advances in cell type experimental access, imaging, optogenetic, and behavioral techniques, such a research program is also increasingly feasible.

Cell Type Genetic Tools

Specific and systematic experimental access to neuronal subpopulations is a prerequisite to exploring circuit organization, function, and assembly (Fig. 1). Over the past decade, my laboratory has made decisive and sustained progress toward a systematic genetic targeting of cortical neuron types. In the first round, we generated and characterized approximately 20 knock-in driver lines that targeted major populations and lineages of GABAergic neurons; these genetic tools have transformed the study of GABAergic circuits. As there is no simple relationship between a single gene and a neuronal type defined by anatomic and physiological features, in a second-round effort we designed multiple combinatorial strategies that engage cell lineage, birth time, and anatomy to substantially improve the specificity of cell type targeting. In the third round since 2013, we have extended these strategies to glutamatergic PyN subpopulations defined by laminar position and projection targets. These tools have enabled us to initiate the project on cortical control of volitional movement (see below). With the recent establishment of a Center for Mouse Brain Cell Atlas that I am leading (U19MH114823-01) in the BRAIN Initiative Cell Census Network (BICCN), we will continue a collaborative project to achieve more comprehensive targeting of forebrain projection neurons, including cortical, striatal, and thalamic projection neurons. These cell type tools will greatly accelerate studying the cortical–striatal–thalamic system, the most prominent network in the mammalian brain, which mediates a wide range of sensory, motor, emotional, and cognitive functions.

Molecular Genetic Basis of Neuronal Identity and Diversity

Understanding the biological basis of neuronal identity (i.e., neuron type)—a group of cells that share anatomical, physiological, molecular, and functional

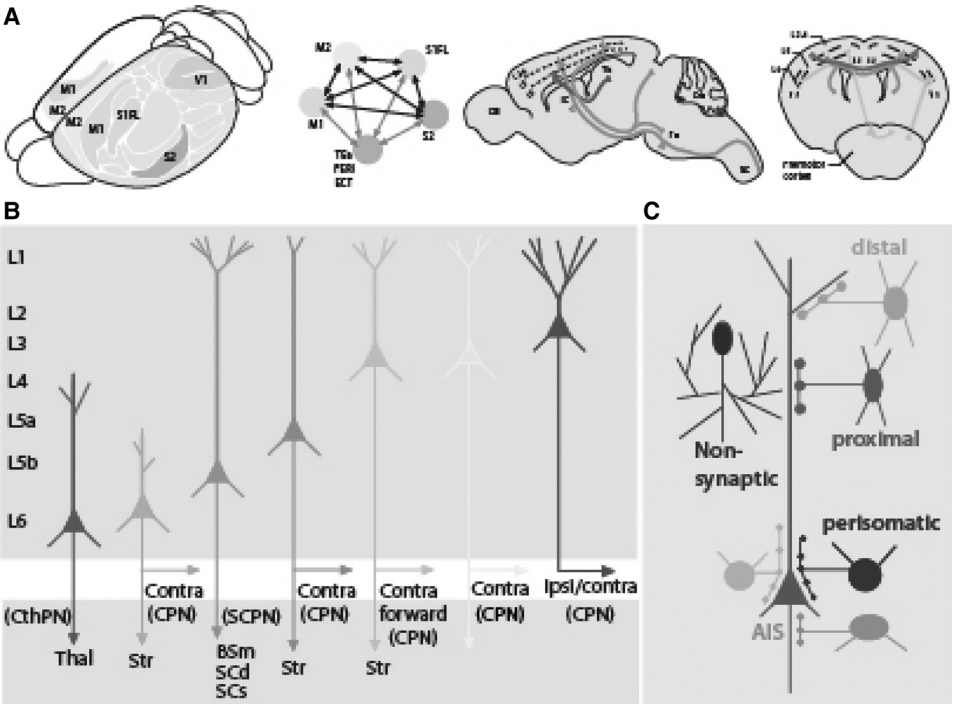


Figure 1. (*A*) Functional cortical areas form processing subnetworks and output channels mediated by pyramidal neurons (PyNs). (*B*) Several major types of PyNs distinguished by laminar position and axon projection. (*C*) Several cardinal types of GABAergic neurons that target different subcellular compartments of PyNs.

properties—is necessary for deciphering neuronal diversity and exploring the organization and assembly of brain circuits. A fundamental conceptual and technical challenge is establishing an overarching and mechanistic framework of neuronal identity that integrates multimodal cell phenotypes. Combining genetic targeting, high-resolution single-cell transcriptomics and computational analysis, we recently discovered that the transcriptional architecture of synaptic communication delineates cortical GABAergic neuron identities (Fig. 2). This architecture comprises six categories of approximately 40 gene families, including cell adhesion molecules, transmitter-modulator receptors, ion channels, signaling proteins, neuropeptides and vesicular release components, and transcription factors. Combinatorial expression of select members across families shapes a multilayered molecular scaffold along the cell membrane that may customize synaptic connectivity patterns and input–output signaling properties. Transcriptional signatures of synaptic communication may integrate anatomical, physiological, functional, and developmental genetic features, which together define neuronal identity. This discovery provides an overarching and mechanistic framework for cell type definition, discovery, and cataloging.

We will continue to use high-resolution molecular profiling in phenotype-defined neuronal subpopulations to examine key issues regarding neuronal identity and diversity by addressing the following questions.

What is the proper granularity and biological basis of neuronal "subtypes"? Within cardinal types, are more "subtle" phenotypic variations best captured as continuum or discrete subgroups? What is the appropriate granularity that reflects biological mechanisms and informs circuit organization? We address these questions through multifaceted molecular, anatomical, developmental, and functional studies, with a particularly robust experimental system of cortical chandelier cells (ChCs; see below).

Does the synaptic communication scheme apply to the definition of other neuron types? With our genetic targeting of PyNs, we are using our high-resolution single-cell RNA–sequencing (scRNA-Seq) method to examine the transcriptional basis of distinct PyN projection classes and subpopulations. I hypothesize that different PyN "types" are also delineated by transcriptional profiles of synaptic communication, likely

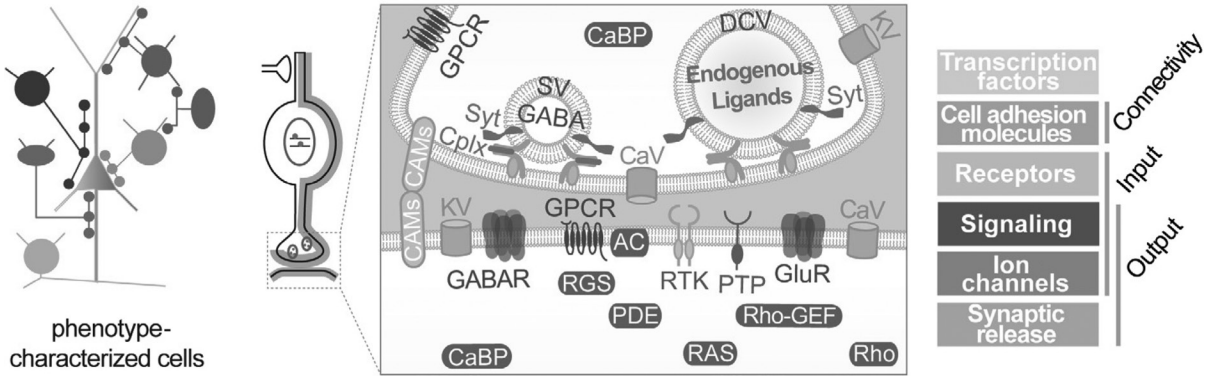


Figure 2. Six cardinal types of GABAergic neurons (*left*) are delineated by their transcription profiles—mainly consisting of six functional gene categories encoding a molecular scaffold that mediates synaptic input–output communication (*right*).

consisting of the same functional gene categories and similar gene families as for GABAergic neurons, but using different combinations of gene family members to shape PyN-characteristic features.

Assembly and Function of a Chandelier-Pyramidal Cell Module

Unlike invertebrate and retinal circuits with readily identifiable connectivity modules (e.g., retinal mosaics), the wiring complexity of cortical networks often precludes the recognition of simpler connectivity motifs and circuit modules. Fortunately, the ChC–PyN connectivity represents an exception and a discrete microcircuit module (Fig. 3). ChCs are the most distinctive GABAergic interneurons that specifically innervate PyNs at the axon initial segment (AIS) and likely control spike initiation. Through genetic fate mapping with the *Nkx2.1-CreER* driver, we discovered the lineage origin of ChCs (Fig. 3). Specified at cell birth, young ChCs are likely endowed with cell-intrinsic programs that guide their long-distance migration and laminar deployment before integration into cortical circuits.

Through collaboration with Qingming Luo (HUST, Wuhan, China), who invented fluorescence micro-optical sectioning tomography (fMOST), which allows axon-resolution and whole-brain light microscopy, we have established a single neuron anatomy platform in the mouse brain. We discovered that ChCs consist of multiple fine-grained subtypes likely delineated by their finer input–output connectivity (e.g., to subsets of PyNs defined by projection targets; Fig. 3). We have further discovered that a subset of layer 2 ChCs mediates highly specific and directional inhibition between two PyN ensembles and cortical subnetworks (Fig. 3). The discovery of bona fide ChC subtypes raises the issues of underlying molecular basis, developmental origin, connectivity pattern, and ultimately functional significance.

Currently we are pursuing the following questions.

What is the developmental origin of ChC laminar subtypes? We have significantly improved the precision of fate-mapping medial ganglionic eminence (MGE) progenitors through intersection of *Nkx2.1-Flp* and *Ascl1-* or *Dlx1-CreER* driver lines to distinguish radial glia (RG) versus several types of intermediate progenitors (IPs). We found that laminar ChC subtypes are, in part, generated as temporal cohorts at different stages of lineage progression distinguished by progenitor types (RGs vs. IPs) and neurogenic mechanisms (direct vs. indirect neurogenesis) (Kelly S, et al., in prep.). Combined with scRNA-seq at different developmental stages, we aim to achieve a more comprehensive understanding of the mechanistic basis of ChC subtypes.

How do ChCs integrate into cortical circuits? On their laminal settlement by the end of the first postnatal week, we found massive apoptosis of young ChCs during the second week that may adjust their density across cortical areas. At the border region between primary (V1) and secondary (V2) visual cortex, ChC density is regulated by the axons and activity of callosal neurons (CNs) of the contralateral visual cortex and retinal activity between P7-P14, before eye

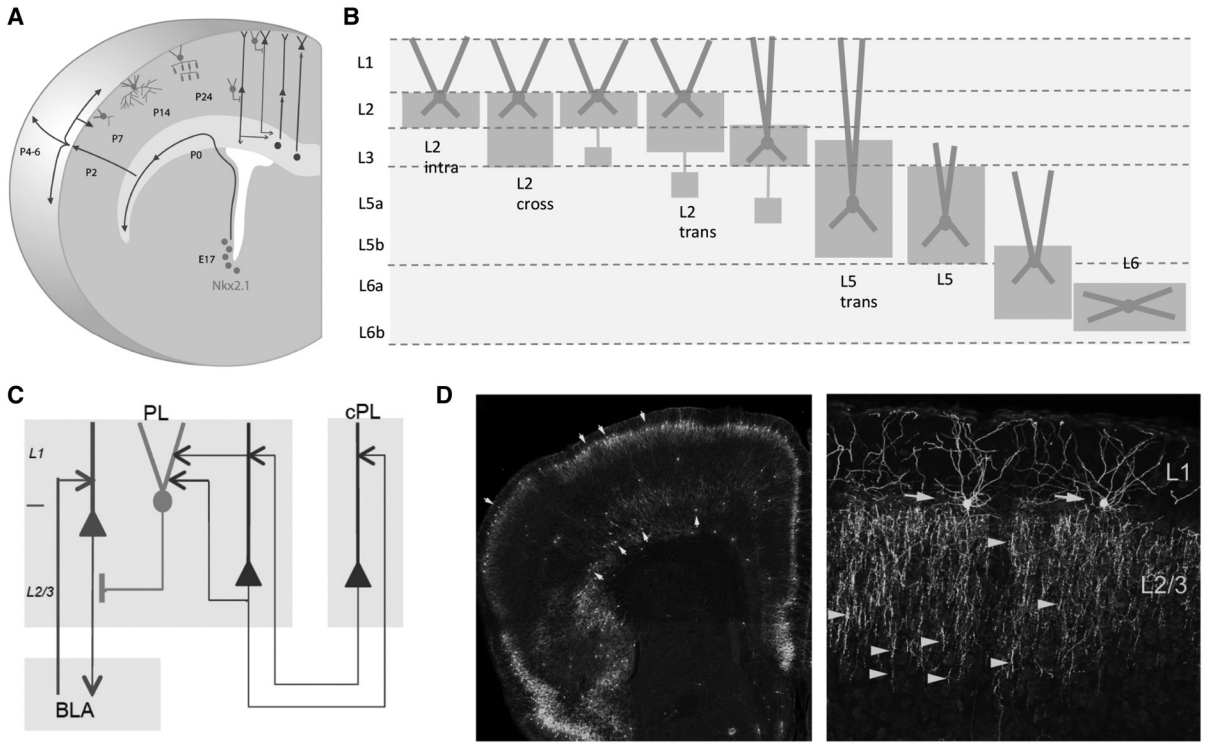


Figure 3. The cortical chandelier cell (ChC)-(pyramidal neuron) PyN module. (*A*) Genetic fate mapping reveals the developmental trajectory of ChCs. (*B*) Single-neuron anatomy reveals multiple fine-grained ChC subtypes likely distinguished by input–output connectivity. (*C*) L2 ChCs mediate directional inhibitory control between PyN ensembles and cortical subnetworks. (*D*) The Unc5b-CreER captures highly specific ChC subsets.

opening. As the V1-V2L border receives inputs from the temporal retina, which represents the central visual meridian that integrates the left and right visual fields, we hypothesize that activity-dependent elimination of ChCs at V1-V2L border might contribute to the development of a fast interhemispheric CN pathway that facilitates an integration of the cortical representation of left and right visual fields (Wang et al., in prep.).

What is the function of ChCs and ChC subtypes in cortical circuit operation? Based on novel markers revealed by scRNA-seq, we have generated two new driver lines that specifically and robustly target ChC subtypes without tedious embryonic manipulations (Fig. 3). These genetic tools enable efficient recording (e.g., GCaMP imaging) and manipulation (optogenetics, DREADD) in the context of specific PyN ensembles and a compelling behavior paradigm (see below). We plan to examine the circuit function of ChCs in the context of cortical control of coordinated movement.

Genetic Dissection of PyN Types Underlying Cortical Output Channels and Processing Streams

We have led a systematic effort to genetically target PyN subpopulations and progenitors. In collaboration with Paola Arlotta (Harvard University), we have generated and characterized more than a dozen Cre and Flp knockin lines targeting PyNs defined by laminar location and axon projection (e.g., Fig. 4), with more focus on L5 and L6 PyNs that constitute multiple output pathways. Combined with antero- and retrograde viral vectors, these tools enable increasingly precise cell type resolution analysis of PyN input–output connectivity, physiological properties, circuit function, and molecular profiles (Matho K, et al., in prep.).

In addition, we have generated inducible and intersectional drivers targeting RG cells (e.g., Lhx2, Fezf2, Pax6), intermediate progenitors (Tbr2 CreER and Flp), and neurogenic progenitors (Tis21). As all PyNs are generated from these progenitors, these tools

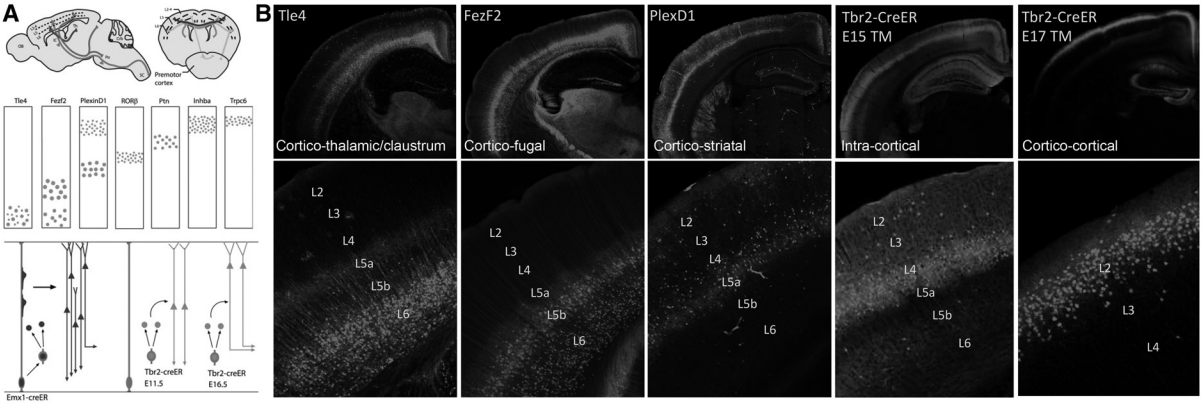


Figure 4. Systematic genetic targeting of cortical pyramidal neurons (PyNs). (*A*) We design genetic targeting strategies for projection classes (*top*) based on laminar expression pattern (*middle*) and cell lineage/birth order mechanism (*bottom*). (*B*) We succeeded in targeting multiple classes cortico-thalamic (Tle4), cortico-fugal (Fezf2), cortico-striatal (PlexD1), and cortico-cortical (Tbr2 embryonic induction) PyNs.

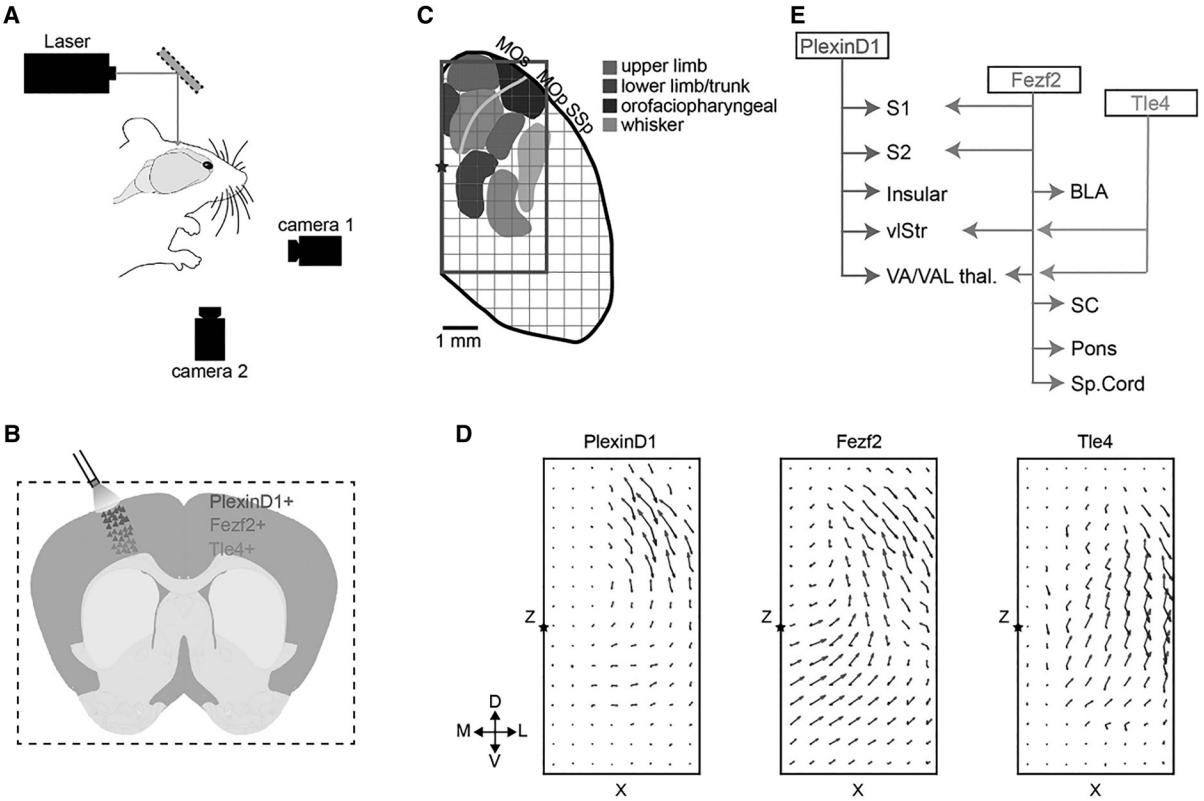


Figure 5. Genetic dissection of pyramidal neuron (PyN) types and circuits underlying cortical control of coordinated movement. (*A*) A simple schematic of optogenetic stimulation and movement monitoring. (*B*) Systematic stimulation of three major classes of PyNs. (*C*) Optogenetic scan in a 3 mm × 6 mm area superimposed on classic sensory and motor areas. (*D*) Forelimb (green) and jaw (red) movement vector maps generated from optogenetic stimulation of three PyN subpopulations in the area indicated in *C*. Note the anterior–lateral area that we call "lateral motor cortex" (LMC; A, 1.5 mm; L, 2.25 mm) where coordinated limb and jaw movements are induced. (*E*) PlexinD1, Fezf2, Tle4 PyNs in LMC appear to form a cortico–striatal–thalamic subnetwork, with Fezf2 PyNs further mediating output to a range of subcortical targets.

begin to enable fate-mapping of PyN types defined by anatomy and connectivity from progenitor types and their lineage progression.

Cortical Circuits Controlling Hand–Mouth Synergy and Feeding

We have initiated a major effort to frame our study of cortical PyN and interneuron types in the functional context of cortical control of voluntary movement. Taking advantage of a comprehensive set of driver lines, we systematically screened the effects of activating major PyN subpopulations in triggering forelimb, jaw, and oral–facial movements in awake, head-fixed mice. Amazingly, we discovered a set of hierarchically organized "action maps" encoded in distinct PyN populations (Fig. 5): Whereas the Fezf2 corticofugal PyNs manifest a topographic movement trajectory map, the Tle4 corticothalamic PyNs feature a similar, yet compressed, map at a shifted location. Most strikingly, activation of PlexinD1 corticostriatal PyNs at a single location triggers highly coordinated bilateral arm, wrist, hand, digit, jaw, and tongue movements that closely resemble self-feeding. Thus, these PlexinD1 PyNs may function as "command neurons" for an innate hand-mouth synergy during feeding. Anterograde, retrograde, and *trans*-synaptic tracing begin to reveal that these three types of PyNs form a cortical and cortico–striatal–thalamic network with pronounced reciprocal connectivity, whereas Fezf2 PyN further project to a range of subcortical targets.

We have now established a head-fixed preparation in which mice engage in robust hand–mouth synergy and feeding. This behavior paradigm allows monitoring and manipulating of distinct PyN types and sets the stage for exploring the underlying circuit mechanisms.

In summary, by establishing a firm grasp of a large group of cortical neuron types and linking their developmental trajectory to circuit function, we are poised to integrate the study of cortical circuit organization and assembly and to reveal the underlying cellular mechanisms.

PUBLICATIONS

Kim Y, Yang GR, Pradhan K, Venkataraju KU, Bota M, García Del Molino LC, Fitzgerald G, Ram K, He M, Levine JM, et al. 2017. Brain-wide maps reveal stereotyped cell-type-based cortical architecture and subcortical sexual dimorphism. *Cell* **171:** 456–469.e22.

Krishnan K, Lau BY, Ewall G, Huang ZJ, Shea SD. 2017. MECP2 regulates cortical plasticity underlying a learned behaviour in adult female mice. *Nat Commun* **8:** 14077.

Lu J, Tucciarone J, Padilla-Coreano N, He M, Gordon JA, Huang ZJ. 2017. Selective inhibitory control of pyramidal neuron ensembles and cortical subnetworks by chandelier cells. *Nat Neurosci* **20:** 1377–1383.

Paul A, Crow M, Raudales R, He M, Gillis J, Huang ZJ. 2017. Transcriptional architecture of synaptic communication delineates GABAergic neuron identity. *Cell* **171:** 522–539.e20.

Vogt D, Cho KKA, Shelton SM, Paul A, Huang ZJ, Sohal VS, Rubenstein JLR. 2017. Mouse *Cntnap2* and human *CNTNAP2* ASD alleles cell autonomously regulate PV$^+$ cortical interneurons. *Cereb Cortex* **28:** 1–12.

Wang X, Tucciarone J, Jiang S, Yin F, Wang B, Wang D, Jia Y, Jia X, Li Y, Yang T, et al. Genetic single neuron anatomy reveals fine granularity of cortical interneuron subtypes. bioRxiv doi: 10.1101/219485.

In Press

Crow M, Paul A, Ballouz S, Huang ZJ, Gillis J. 2018. Characterizing the replicability of cell types defined by single cell RNA-sequencing data using MetaNeighbor. *Nat Commun* **9:** 884.

He M, Huang ZJ. 2018. Genetic approaches to access cell types in mammalian nervous systems. *Curr Opin Neurobiol* **50:** 109–118.

NEUROBIOLOGY OF COGNITION AND DECISION-MAKING

A. Kepecs M. Bosc, Q. Chevy, E. Demir, E. Gibson, T.S. Gouvea, Q.A. Li, S. Li, P. Masset, T. Ott, H.J. Pi, T. Pinkhasov, L. Ramirez Sanchez, S. Ren, S. Starosta, J.F. Sturgill

Our long-term goal is to "reverse engineer" the computational and neurobiological principles underlying cognition and decision-making and apply these insights to biological psychiatry. Neuroscience has made great strides toward understanding systems for sensory processing and motor output. Yet, studying cognition at the level of neurobiological substrate seemed hopeless until recently. The goal of our laboratory is to bridge the chasm between cognitive and circuit neuroscience by applying a unique, multifaceted skill set and approach. We translate psychological questions into the language of neuroscience by developing quantifiable, well controlled behavioral tasks for rodents. We then couple these tasks with targeted, high-resolution and high-throughput monitoring and manipulation of the neural circuits mediating cognitive behaviors. Given the complexity of animal behavior and the dynamics of the neural networks that produce it, our studies rely on computational models to guide and sharpen the neurobiological questions. Finally, we use human psychophysics to validate our behavioral observations in rodents by linking them with analogous behaviors in humans. Behavioral links to humans can also serve to identify behavioral dimensions that are predictive of disordered mental states, and our goal is to bridge our studies in animals to psychiatric disorders.

In terms of topics, our approach is multifaceted: We study (i) the roles of uncertainty and confidence in decision-making, (ii) foraging decisions about whether to stay or to switch, (iii) the division of labor between different cell types in prefrontal cortex, (iv) how the cholinergic system supports learning, and (v) social decisions that rely on shallow, stereotyped circuits. A unifying theme is the use of cell type– and pathway-specific perturbations to effect gain and loss of function for specific behavioral abilities. We hope that by identifying the neural processes underlying behavior in "humanized" mouse models of cognition, we can understand what goes awry in the brain during mental illness. Ultimately we hope these insights will enable us to develop novel therapeutic strategies for psychiatric disorders such as addiction, major depression, schizophrenia, and autism spectrum disorder.

Confidence in Rats, Humans, Brains, and Statistics

P. Masset, T. Ott, M. Bosc, T.S. Gouvea, S. Ren

Every decision we make is associated with a feeling of confidence. The long-term goal of our research is to understand how the brain implements confidence judgments and acts on these. The appropriate estimates of confidence can drive information-seeking behaviors, learning, and attention by reducing the level of uncertainty. Conversely, the pathological misevaluation of confidence contributes to a wide range of neuropsychiatric conditions, including anxiety, obsessive-compulsive disorder, and addiction. We have developed a set of behavioral tasks and a theoretical framework that rigorously translates the psychological concept of confidence into a formally defined decision variable. Using this approach, we identified orbitofrontal cortex (OFC) neurons that encode the confidence associated with a perceptual decision. Further, we have derived a mathematical framework for decision confidence from first principles of statistics (Hangya et al. 2016). We showed that key properties of statistical decision confidence match human self-reported confidence (Sanders et al. 2016), providing a deep link between objective and subjective notions of confidence. We also showed that reversible lesions of OFC specifically disrupt confidence reporting (Lak et al., *Neuron 84:* 190 [2014]).

OFC has been previously shown to play a central role in the confidence estimate, but it remains unknown whether OFC locally computes confidence or receives confidence signals from other areas. Choice

and confidence could be computed together in sensory areas and then relayed to OFC. To test this hypothesis, we are studying the impact of auditory cortex lesions on auditory discrimination and its subsequent confidence estimate in rats performing a postdecision temporal wagering task.

Building a confidence representation requires two steps: estimating the uncertainty of the information, such as the reliability of a sensory perception, and then predicting the probability of an outcome given the information. To test to which level in this process the confidence representations in OFC can be attributed, we recorded single neurons in OFC during a dual sensory modality decision task with a confidence report. Our results reveal that single neurons in OFC can represent decision confidence irrespective of the sensory modality used to make the decision. Furthermore, the firing rates of these neurons predict the behavioral report of decision confidence. Therefore OFC appears to contain a modality-general representation of confidence that could provide an information source–independent probability estimate useful for confidence-driven adaptive behaviors, such as learning and time investment.

Beyond our rodent work, we are now in a position to use these quantitative measures of decision confidence in humans, through collaborations, with neuroimaging and genetic approaches. Moreover, by showing that a single confidence measure is applicable to humans and rodents, our results strengthen the case for using the rat as a model system for translational studies in cognition and psychiatry.

Categorical and Cell Type–Specific Representations of Decision Variables in Orbitofrontal Cortex

This work was done in collaboration with J. Hirokawa, Doshisha University; A. Vaughan

OFC, like other cortical regions, creates internal representations of the external world in the form of neural activity, which is structured to support adaptive behavior. In many cortical regions, individual neurons respond to specific features that are matched to the function of each region and to statistics of the world. In OFC, like elsewhere in frontal cortex, neurons display baffling complexity, responding to a mixture of sensory, motor, and other variables. Here we use an integrated approach to understand the architecture of higher-order cortical representations and show that discrete groups of OFC neurons encode distinct decision variables. Using rats engaged in a complex task combining perceptual and value-guided decisions, we found that OFC neurons can be grouped into distinct, categorical response types. These categorical representations map directly onto decision variables, such as reward size, decision confidence, and integrated value, in a choice model explaining our task. We propose that, like sensory neurons, frontal neurons form a sparse and overcomplete population representation aligned to the natural statistics of the world—in this case spanning the space of decision variables required for optimal behavior.

Neural Circuit Logic of Orbitofrontal Cortex during a Postdecision Confidence Task

T. Ott, S. Ren, P. Masset [in collaboration with J. Hirokawa, Doshisha University; M. Lagler, T. Klausberger, Medizinische Universität Wien]

To understand specific neuron types within OFC we use two complementary approaches. We first target specific projection neurons using retrograde viruses and use optogenetic stimulation to identify these in electrophysiological recordings. Using this technique, we have found that OFC projections to ventral striatum show a characteristic response in which negative value signals are sustained throughout the intertrial interval to the beginning of the subsequent trial. Second, in collaboration with the Klausberger laboratory, we use juxtacellular labeling to target neurons based on their functional response profiles. Once labeled, neurons are subjected to detailed ex vivo analysis of their axonal projection patterns. Using these techniques we have begun to record and identify OFC neurons that specifically signal the confidence-dependent waiting time. We expect that the combination of juxtacellular and optogenetically identified extracellular recordings will enable us to reverse engineer the cell type–specific circuit logic of orbitofrontal cortex.

Confidence-Dependent Updating of Choice Strategy during Perceptual Decisions

P. Masset, T. Ott [in collaboration with A. Lak, University of Cambridge]

Past experiences are invaluable to improve future decisions. Learning from past rewards (i.e., reinforcement learning) ensures adaptive decisions that yield

the largest possible rewards. To what extent are the reinforcement learning mechanisms at work during decisions that are informed by available sensory signals rather than past rewards? We identified a novel form of reinforcement learning during perceptual decisions that depends on the uncertainty of past sensory judgments. We show that past rewards influence perceptual decisions mainly when the previous sensory stimulus was difficult to judge and thus the confidence in obtaining the reward was low. This choice updating was independent of the modality of sensory stimulus based on which the previous choice was made: The updating was transferred from olfactory to auditory decisions, and vice versa. We are investigating which reinforcement learning models can account for this form of learning. Our results reveal that learning from previous choices is strongly modulated by the confidence associated with those decisions. They demonstrate that animals and humans continually incorporate reward feedback into their choice strategies even during well trained perceptual decisions.

Neural Circuits and Quantitative Measures of Impulsive Choice in Mice

T. Pinkhasov, S. Starosta, H.J. Pi

Impulsivity is a behavioral trait present in many psychiatric disorders that significantly increases the risk of suicide, violence, and criminal behavior. Whether a particular decision made too early is impulsive is difficult to determine, because misevaluation of expected outcome or misestimation of time could lead to similar consequences. Therefore, as an initial step toward understanding the underlying neural circuits, we sought to develop a task that isolates the contribution of impulsivity to individual choices and separates it from reward valuation. We devised a novel behavioral paradigm that enables us to parse out the contribution of these pathways to reward valuation and motor inhibition. Our goal is to elucidate the contribution of neural circuits that are involved in impulsivity, particularly the anterior cingulate cortex's control of the serotonergic and dopaminergic systems, using photometry recordings and optogenetic manipulation. Ultimately, we hope our circuit-based understanding of impulsivity will contribute to the design of circuit-specific treatments for impulsivity disorders.

Circuit and Behavioral Functions of Cortical Chandelier Neurons

Q.A. Li [in collaboration with Z.J. Huang, J. Tucciarone, CSHL]

Chandelier cells (ChC) are perhaps the most unique GABAergic interneurons in cortex. They specialize in innervating the axon initial segment of excitatory pyramidal neurons, the site for action potential generation—yet it remains unclear whether they function to inhibit their targets. Taking advantage of a genetic and viral approach developed in the Huang lab, we are able to target a subgroup of ChCs in PFC. By combining electrical stimulation at the basolateral amygdala (BLA) and extracellular recordings in PFC, we are aiming to determine whether ChCs inhibit or excite their BLA-projecting pyramidal neurons. We are also developing a behavioral paradigm for testing approach–avoidance conflict in head-fixed mice that engage these circuits. We aim to identify ChCs' impact on their local circuit and evaluate when they are recruited during behavior.

Cortical VIP Interneurons, Disinhibitory Control, and Reinforcement Learning

Q. Chevy, H.J. Pi [in collaboration with Z.J. Huang, CSHL; Z. Szadai, B. Rozsa, Hungarian Academy of Sciences]

We have recently identified a disinhibitory cortical circuit motif that appears to be a conduit for fast neuromodulatory action in cortex. This circuit is controlled by a class of inhibitory interneurons that express vasoactive intestinal polypeptide (VIP) and inhibit other interneurons, thereby disinhibiting a subpopulation of principal neurons. Functionally, we showed that VIP interneurons in the auditory cortex are recruited in response to specific reinforcement signals such as reward and punishment.

To explore the generality of these observations across cortex regions, we are collaborating with B. Rozsa (KOKI, Hungarian Academy of Sciences) in the use of a state-of-the-art 3D random-access AOD two-photon imaging system to record the sparse VIP population across large regions. We found that most VIP neurons are activated by reward and punishment across multiple cortical regions, suggesting that their behavioral recruitment has a cortex-wide function in reinforcement learning.

We have also begun to record VIP neurons in a cued-outcome task to determine what aspects of reinforcement they signal. Do they respond simply to the delivery of primary reinforcers, reinforcement prediction errors, or reinforcement surprise? We found that cue responses emerge with learning and are proportional to outcome value, suggesting a more sophisticated role in reinforcement learning. By establishing the detailed circumstances under which VIP interneurons are recruited and identifying their generality and circuit mechanisms, we expect that these studies will reveal fundamental principles about cortical microcircuits that are applicable across cortical regions.

Neural Representation of Social Decisions and Rewards

E. Demir, L. Ramirez Sanchez [in collaboration with R. Axel, Columbia University; N. Bobrowski-Khoury]

Social behavior is integral to animals' survival and reproduction; social deficits are at the heart of cognitive disorders such as autism spectrum disorder that have proven profoundly difficult to study in model organisms. We would like to understand how social information is represented, computed, and used by mice. In rodents, a main source of information for social decision-making and reward valuation is the chemosensory system.

One component of these scents is darcin, a mouse urinary protein that is sufficient to induce innate attraction in sexually receptive female mice and also acts as an unconditioned stimulus in associative learning paradigms. We have identified a neural circuit that extends from the vomeronasal organ to the medial amygdala and mediates the innate response to darcin. Genetic silencing of either accessory olfactory bulb or the medial amygdala eliminates the innate attraction to darcin. Optical reactivation of darcin-activated medial amygdala neurons elicits attraction behavior.

We are also interested in understanding basic rules that mice use to choose partners. For this purpose, we have developed a psychophysical social behavior task, the "social carousel," inspired by perceptual psychophysics and game theoretic traditions that have been instrumental to studying other facets of cognition. Our task enables reliable, quantitative, and high-throughput analysis of social interactions in mice. Subjects can choose to engage in extended social interactions with the caged mice at the expense of delaying the water reward. This task allows us to infer the "social value" of a mouse based on the trade-off between social interactions and appetitive rewards. In addition, this task is compatible with our electrophysiological studies, because it is devised for precise stimulus delivery and reproducible behavioral contingencies.

Basal Forebrain Cholinergic and GABAergic Neurons in Attention and Learning

J.F. Sturgill, S. Li [in collaboration with B. Hangya, Hungarian Academy of Sciences]

Basal forebrain (BF) cholinergic neurons constitute a major neuromodulatory system implicated in normal cognition functions, including learning, memory, and attention. Cognitive deficits in Alzheimer's disease, Parkinson's dementia, age-related dementias, and normal aging are correlated with the extent of deterioration of BF cholinergic neurons. Cholinergic projections densely innervate neocortex and release acetylcholine, which is thought to regulate arousal, attention, and learning.

We recorded optogenetically identified cholinergic neurons using optogenetic identification in mice performing an auditory detection task requiring sustained attention. We found that cholinergic neurons responded to reward and punishment. This reinforcement response invites comparison to dopamine neurons, for which a key conceptual advancement was that they compute reward prediction error—the difference between reward expectation (as informed by a predictive cue) and the reward received. Therefore, we used fiber photometry to measure bulk GCaMP signal in cholinergic neurons in a cued probabilistic outcome task. After the mice learned the task, cholinergic neurons responded to reward-predicting cues, and the neurons' reward responses were diminished by cued expectations, similar to dopamine neurons. These results reveal that the cholinergic system broadcasts a rapid and precisely timed reinforcement signal that could support fast cortical activation and plasticity. The basal forebrain also contains long-range GABAergic projections to the cortex, but this is poorly understood. Using optogenetic identification of PV^+ NB neurons, we found that, unlike cholinergic neurons, these neurons encode expected value of the cue and predict the reaction time of mice on a trial-to-trial

basis. We expect that these studies will reveal fundamental principles about how the parallel GABAergic and cholinergic projection subsystems in the basal forebrain differentially control cortical processing.

A Receptor Complementation Strategy for Efficient, Tropism-Free Retrograde Targeting of Neurons

S. Li, J.F. Sturgill

Retrogradely transported neurotropic viruses enable researchers to target neurons based on their long-range projections and have thus become indispensable tools for linking neural connectivity with function. A major limitation of viral techniques is that they rely on cell type–specific molecules for uptake and transport and, as a result, may fail to infect neurons that do not express the requisite complement of surface receptors (viral tropism). To overcome this problem, we designed AAV constructs to express the coxsackievirus and adenovirus receptor (CAR) and thereby potentiate canine adenovirus type 2 (CAV-2) infection in candidate projection neurons. Enhancement of CAR expression greatly increased retrograde labeling rates with CAV-2 in multiple long-range projecting neural circuits in both mice and rats, providing a robust method for high-efficiency, tropism-free retrograde labeling. The CAR/CAV-2 system will facilitate efficient retrograde targeting for functional analysis of neural circuits.

Nanophotonic Silicon Probes for Multisite Optical Stimulation

Q.A. Li [in collaboration with A. Mohanty, M.A. Tadayon, and M. Lipson, Columbia University]

The ability to activate neural populations using optogenetics has revolutionized the study of neural circuits. However, this is currently primarily done by using a single fiber to flood light into a large volume of the brain. We demonstrate the first active optical switch for individual neuron excitation at 473 nm, which enables one to control multiple independent beams for deep brain neural stimulation. Using a silicon nitride waveguide platform, the Lipson group developed a cascaded Mach-Zehnder interferometer (MZI) network located outside the brain to direct light to eight different grating emitters located at the tip of the neural probe. To demonstrate the potential of the platform and provide bidirectional neural interfacing for localized and specific stimulation and recording in deep brain area in vivo, we package the device with tungsten wires placed near the output gratings with a resolution of <10 microns. We report fast (up to 200 Hz), specific, and independent control of single neuron activity across cortical laminar and hippocampus in anesthetized mice. The demonstrated depth-specific manipulation approaching single-cell resolution shows that this device can control neural activity in vivo independently across beams and with high spatiotemporal resolution.

PUBLICATIONS

Hirokawa J, Vaughan A, Kepecs A. 2017. Categorical representations of decision-variables in orbitofrontal cortex. bioRxiv doi: 10.1101/135707

Lak A, Nomoto K, Keramati M, Sakagami M, Kepecs A. 2017. Midbrain dopamine neurons signal belief in choice accuracy during a perceptual decision. *Curr Biol* **27:** 821–832.

Mohanty A, Li Q, Tadayon MA, Bhatt GR, Cardenas J, Miller SA, Kepecs A, Lipson M. 2017. Reconfigurable visible nanophotonic switch for optogenetic applications. *Proc SPIE* 10052, doi:10.1117/12.2251663.

In Press

Chevy Q, Kepecs A. 2018. When acetylcholine unlocks feedback inhibition in cortex. *Neuron* **97:** 481–484.

Li SJ, Vaughan A, Sturgill JF, Kepecs A. 2018. A viral receptor complementation strategy to overcome CAV-2 tropism for efficient retrograde targeting of neurons. *Neuron* **98:** 905–917.e5.

Mohanty A, Li Q, Tadayon MA, Bhatt GR, Cardenas J, Miller SA, Kepecs A, Lipson M. 2018. A reconfigurable nanophotonics platform for sub-millisecond, deep brain neural stimulation. arXiv:1805.11663.

MATHEMATICAL PRINCIPLES OF NEURAL COMPUTATION

A. Koulakov B. Baserdem H. Giaffar D. Kepple S. Shuvaev T. Tran

Our laboratory works on theories of neural computation. Our overall strategy is to use methods developed in mathematics, physics, machine learning, computer science, and statistics to build experimentally testable models of neural networks and their function. In most cases, we base our theories on what is known about particular biological systems; however, given that the principles of brain function remain unclear, in many cases, we resort to building machine learning theories. This means that we formulate the problems solved by the brain in a mathematically rigorous fashion and hypothesize how an engineer would solve the problem. We then use these solutions to form experimentally testable predictions. Testing these predictions in collaboration with our experimental colleagues helps us refute or refine our theories. For example, we are interested in understanding how connectivity is established in the brain. We have proposed several theories that may determine the rules of making connections between neurons based on a limited set of instructions contained in the genome. These theories address several levels of organization, including computational, biological, engineering, and evolutionary. Our theories may explain the differences between connectivities in normal and abnormal brain circuits. We are also interested in understanding the principles of perceptual invariance—that is, how can sensory systems represent objects in the environment despite substantial variations in intensity and background. Visual percepts, for example, retain basic features, such as perceived shape and color composition, despite variable luminance, spectral composition, scale, and position of the stimuli. Although we study the question of perceptual invariance in application to well defined problems, we believe that the principles that we will uncover may generalize across sensory modalities. Finally, we are pursuing the question of how modern theories of machine learning and artificial intelligence can apply to brain function. Although reinforcement learning, deep learning, long short-term memory networks, etc., are successful in solving a variety of artificial intelligence problems, their mapping onto brain circuits remains unclear. We attempt to bring these systems closer to satisfying the constraints imposed by biology. We hope that the convergence of machine learning theories and biology will help us learn more about brain function.

Representations of Sound in Deep Learning of Audio Features from Music

S. Shuvaev, H. Giaffar, A. Koulakov

The work of a single musician, group, or composer can vary widely in terms of musical style. Indeed, different stylistic elements, from performance medium and rhythm to harmony and texture, are typically exploited and developed across an artist's lifetime. Yet, there is often a discernible character to the work of, for instance, individual composers at the perceptual level—an experienced listener can often pick up on subtle clues in the music to identify the composer or performer. Here we suggest that a convolutional network may learn these subtle clues or features given an appropriate representation of the music. In this work, we apply a deep convolutional neural network to a large audio data set and empirically evaluate its performance on audio classification tasks (see Fig. 1). Our trained network demonstrates accurate performance on such classification tasks when presented with 5-sec examples of music obtained by simple transformations of the raw audio waveform. A particularly interesting example is the spectral representation of music obtained by application of a logarithmically spaced filter bank, mirroring the early stages of auditory signal transduction in mammals. The most successful representation of music to facilitate discrimination was obtained via a random matrix transform (RMT). Networks based on logarithmic filter banks and RMT were able to correctly guess the one composer out of 31 possibilities in 68 and 84 percent of cases, respectively (Shuvaev et al. 2017b).

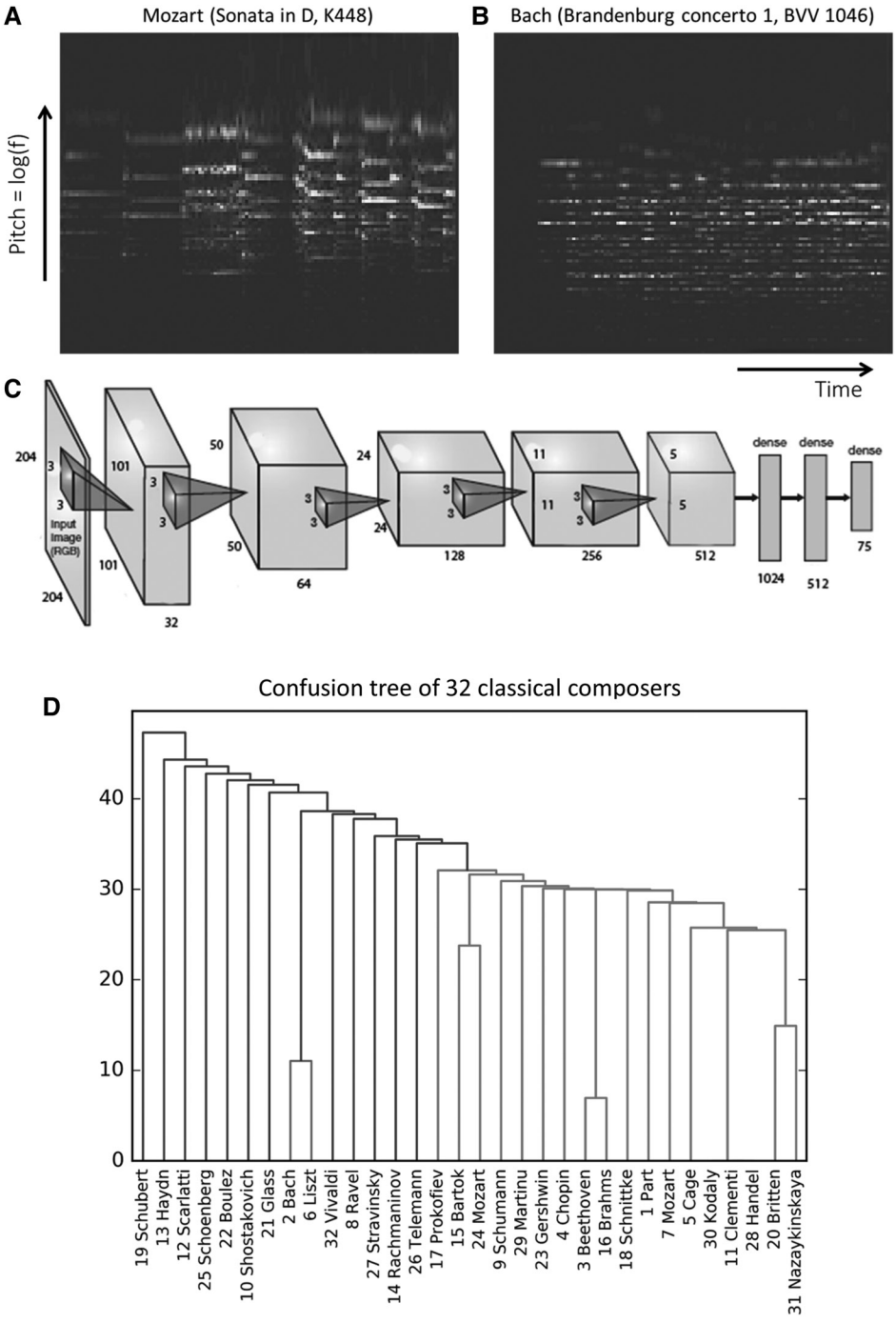


Figure 1. Training neural network to recognize classical music composers. Building such a network creates an analog of the auditory system. (*A*,*B*) Inputs into neural networks are defined as spectrograms. (*C*) The architecture of the deep neural network trained to call the composers. (*D*) The dendrograms of phylogenetic relationships between composers based on network's confusion matrix.

Constructing an Olfactory Perceptual Space and Predicting Percepts from Molecular Structure

D. Kepple, A. Koulakov

Given the structure of a novel molecule, there is still no one who can reliably predict what odor percept that molecule will evoke. The challenge comes both from the difficulty in quantitatively characterizing molecular structure and from the inadequacy of language to fully characterize olfactory perception. Here, we present a novel approach to both problems. First, we avoid explicit characterization of molecular structure by using a similarity score for each molecular pair, derived from comparing the molecular structures directly. We show that this method is advantageous to conventional methods, improving predictions without relying on preexisting knowledge of chemical descriptors. Second, we generate a perceptual space in which a molecule's location defines its percept. We show that from a molecule's neighbors in this space alone, we are able to reproduce all perceptual descriptors of that molecule. We propose that predicting olfactory percept from structure can be rethought of as predicting a molecule's location in this perceptual space. This suggestion provides a framework for understanding and predicting human smell percepts (Kepple and Koulakov 2017).

DALMATIAN: An Algorithm for Automatic Cell Detection and Counting in 3D

S. Shuvaev, A. Koulakov [in collaboration with G.N. Enikolopov, SUN Stony Brook]

Current 3D imaging methods, including optical projection tomography, light-sheet microscopy, block-face imaging, and serial two photon tomography, enable visualization of large samples of biological tissue. Large volumes of data obtained at high resolution require development of automatic image processing techniques, such as algorithms for automatic cell detection or, more generally, point-like object detection. Current approaches to automated cell detection suffer from difficulties originating from detection of particular cell types, cell populations of different brightness, and nonuniformly stained and overlapping cells. In

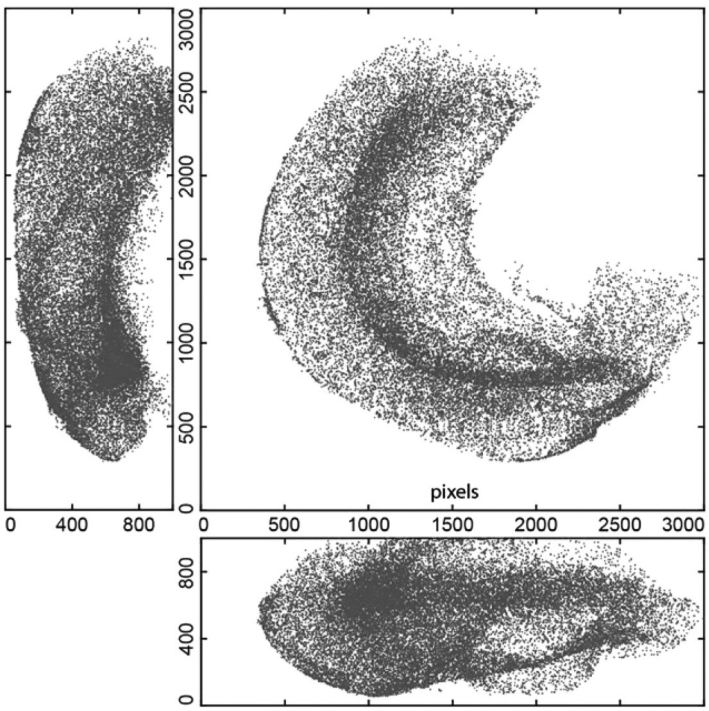


Figure 2. All EdU+ cells detected by a fully automatic computational algorithm in whole-mount stained 3D hippocampus of P14 mouse (three projections). This reconstruction includes 36,451 cells.

this study, we present a set of algorithms for robust automatic cell detection in 3D. Our algorithms are suitable for, but not limited to, whole brain regions and individual brain sections. We used a watershed procedure to split regional maxima representing overlapping cells. We developed a bootstrap Gaussian fit procedure to evaluate the statistical significance of detected cells. We compared cell detection quality of our algorithm and other software using 42 samples, representing six staining and imaging techniques. The results provided by our algorithm matched manual expert quantification with signal-to-noise-dependent confidence, including samples with cells of different brightness and overlapping cells for whole brain regions and individual tissue sections. Our algorithm provided the best cell detection quality among tested free and commercial software (see Fig. 2).

PUBLICATIONS

Kepple D, Koulakov A. 2017. Constructing an olfactory perceptual space and predicting percepts from molecular structure. arXiv:1708.05774.

Kepple DR, Giaffar H, Rinberg D, Koulakov AA. 2017. Primacy coding in dual olfactory networks. In *51st Asilomar Conference on Signals, Systems, and Computers*, 2017-O. pp. 587–592. ASMS, Santa Fe, New Mexico.

Shuvaev SA, Lazutkin AA, Kedrov AV, Anokhin KV, Enikolopov GN, Koulakov AA. 2017a. DALMATIAN: an algorithm for automatic cell detection and counting in 3D. *Front Neuroanat* **11:** 117.

Shuvaev SA, Giaffar H, Koulakov AA. 2017b. representations of sound in deep learning pf audio features from music. arXiv:1712.02898.

THE FUNCTION AND PLASTICITY OF CENTRAL SYNAPSES IN ADAPTIVE AND MALADAPTIVE BEHAVIORS RELATED TO PSYCHIATRIC DISORDERS

B. Li S. Ahrens, A. Bouhuis, C. Bravo-Rivera, H. Deng, C. Fernandes Henriques, A. Furlan, J. Giovanniello, O. Gschwend, W. Guan, L. Mejia, D. Rebolini, X. Xiao, T. Yang, K. Yu, X. Zhang

Understanding the relationship among synapse, circuit, and behavior has been the focus of research in my lab. We are particularly interested in understanding the synaptic and circuit mechanisms underlying cognitive functions, as well as synaptic and circuit dysfunction that may underlie mental disorders, including anxiety, depression, schizophrenia, and autism. To address these questions we use in vitro and in vivo electrophysiology, imaging, molecular, genetic, optogenetic, and chemogenetic methodologies to probe and manipulate the function of specific neural circuits in the rodent brain and determine their roles in behaviors, including fear regulation, anxiety, reward and punishment processing, and habitual and repetitive behaviors related to autism. We are currently undertaking the following major lines of research.

The Role of the Amygdala Circuitry in Fear Regulation and Anxiety

Our previous studies show that the central amygdala (CeA) has a key role in learning and expression of defensive responses to threats. In particular, our studies indicate that somatostatin-expressing (SOM^+) neurons in the lateral division of the central amygdala (CeL) are essential for the acquisition and recall of conditioned freezing behavior—which has been used as an index of defensive response in laboratory animals—during Pavlovian fear conditioning. We also show that SOM^+ CeL neurons are activated by threat-predicting sensory cues following fear conditioning, and that activation of these neurons suppresses ongoing actions and converts an active defensive behavior to a passive response. Furthermore, inhibition of these neurons using optogenetic or molecular methods promotes active defensive behaviors. Our results provide the first in vivo evidence that SOM^+ neurons represent a CeL population that acquires learning-dependent sensory responsiveness during fear conditioning, and furthermore reveals an important role of these neurons in gating passive versus active defensive behaviors in animals confronted with threat.

More recently, we show that another major class of CeL neurons, the protein kinase C-δ-expressing (PKC-δ^+) neurons, is essential for the synaptic plasticity underlying learning in the lateral amygdala, as it is required for lateral amygdala neurons to respond to unconditioned stimulus (US) and, furthermore, carries information about the US to instruct learning. Our results indicate that PKC-δ^+ CeL neurons constitute a key node in a pathway that imparts information about US to the LA during fear conditioning, hence revealing a previously unknown amygdala functional organization in which the CeL is upstream of the lateral amygdala in processing aversive US during learning. Our findings also revise a prevalent model for the functional organization of amygdala circuits, which posits that PKC-δ^+ CeL neurons are "fear-off" neurons—a CeL population that shows inhibitory conditioned responses following fear conditioning—that act to suppress fear responses through inhibition of amygdala output. In fact, we show that a substantial population of PKC-δ^+ CeL neurons are essentially "fear-on" neurons and function in the opposite manner by conveying aversive US signals (Fig. 1; Yu et al. 2017).

The Basal Ganglia Circuit in Motivated Behaviors

The basal ganglia, a group of subcortical nuclei, play a crucial role in motivated behaviors. Recently, we show that neurons in the habenula-projecting globus pallidus (GPh), an output of the basal ganglia, are essential

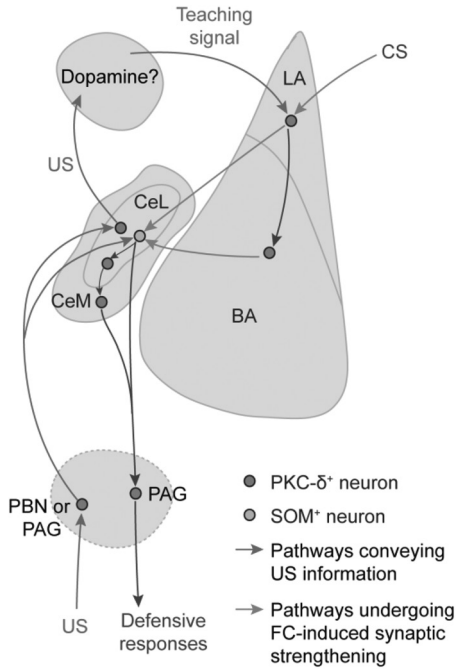


Figure 1. A revised circuit model for classical fear conditioning. For clarity, only key known components of the circuitry are shown. The unconditioned stimulus (US) signals likely come from the parabrachial nucleus (PBN) or the periaqueductal gray (PAG), in which neurons have been shown to provide either nociceptive or teaching signals, respectively, to the amygdala during fear conditioning. Interestingly, it has been shown that midbrain dopamine neurons, including those in the SNc, play an important role in fear learning, and that CeA neurons preferentially innervate GABAergic neurons over dopamine neurons in the SNc/VTA. Therefore, it is possible that PKC-δ+ CeL neurons drive disinhibition of dopamine neurons in response to US, thereby instructing learning in the LA (as indicated by the question mark). Alternatively, or additionally, neurons in the SI or the RRF may also mediate the function of PKC-δ+ CeL neurons. The divergent projections of these neurons also suggest that PKC-δ+ CeL neurons are functionally heterogeneous. Indeed, previous studies indicate that these neurons regulate feeding. The SOM+ CeL neurons, on the other hand, drive freezing behavior during fear conditioning through either the CeM or their long-range direct projections to the PAG. BA, Basal amygdala; CeL, central amygdala, lateral part; CeM, central amygdala, medial part; LA, lateral amygdala; RRF, retrorubral field; SI, substantia innominate; SNc, substantia nigra, compact part; VTA: ventral tegmental area.

for evaluating action outcomes. Our current study addresses the roles of another basal ganglia output, the ventral pallidum, in motivated behaviors.

Motivated behaviors can be driven by two opposing processes, the desire to obtain a reward or the drive to escape a punishment. The ventral pallidum is critical for attributing motivation salience to cues that predict reward and invigorating reward-seeking behavior. However, sparse evidence suggests that it may also play a role in motivating avoidance behavior. We found that two genetically distinct ventral pallidal populations encode the motivation to approach or avoid. Single-unit recording in the ventral pallidum of mice undergoing conditioned reward and punishment tasks revealed three functionally distinct clusters, which encoded (1) incentive salience, (2) aversive salience, and (3) general vigor. Optogenetic inhibitory tagging showed that neurons that encoded incentive salience were GABAergic, as were the neurons that encoded general vigor. In a Pavlovian reward task, these incentive salience neurons encode the expected reward value through a phasic increase in neuronal firing. As with dopamine neurons, the response to expected reward diminishes with training; but unlike dopamine neurons, the response to reward cues inverts and is signaled with a decrease in neuronal firing when mice are sated. This suggests that GABAergic ventral pallidal neurons encode a state-dependent prediction error signal. These neurons also signal state value on a longer timescale, as their baseline firing was modulated by the state depending on block type (reward vs. punishment). Taken together, GABAergic ventral pallidum (VP) neurons signal the positive motivational state and incentive value to invigorate reward seeking on a short and long timescale. In contrast, the glutamatergic ventral pallidal neurons were phasically excited by the expectation or delivery of punishment. As with GABAergic neurons, the baseline firing was modulated depending on the block condition, indicating that these neurons encode aversive salience on both long and short timescales. Together our data show that there are two genetically distinct types of ventral pallidal neurons that encode either incentive or aversive salience and are critical for driving reward approach or punishment avoidance.

Circuit Mechanisms of the Distinct Phenotypes of a 16p11.2 Deletion Model of Autism

Previous studies and preliminary results in our lab indicate that mice carrying a deletion corresponding to the human 16p11.2 microdeletion (the "16p11.2 mice") have behavioral changes, including (1) increased locomotor activity and repetitive behaviors, (2) cognitive deficit, and (3) enhanced anxiety. These

behavioral changes, which are related to the symptoms of humans carrying the homologous deletion, suggest potential dysfunctions in the basal ganglia, a group of nuclei in the brain that have been implicated in motor function and cognition, as well as emotion- and anxiety-related behaviors. Notably, recent studies have identified cellular and structural changes in the basal ganglia of the 16p11.2 mice, which may underlie both the motor and cognitive deficits in these mice. Furthermore, the changes in the basal ganglia may also contribute to the anxiety phenotypes of these mice.

On the basis of all these findings, including our own, we are currently testing the hypothesis that impaired function in different nuclei of the basal ganglia causes the divergent behavioral changes in the 16p11.2 mice, including cognitive deficit, increased locomotor activity, and enhanced anxiety. We are testing this hypothesis using a combination of state-of-the-art neuroscience technologies, including electrophysiology, imaging, optogenetics, chemogenetics, and novel behavioral techniques. This line of research, which addresses the roles of the basal ganglia in distinct behavioral changes in the 16p11.2 deficiency model of autism, will establish a framework for research into the mechanisms underlying the core and comorbid symptoms of autism spectrum disorders, and will guide the development of novel and effective therapeutics.

PUBLICATIONS

Bravo-Rivera C, Stephenson-Jones M, Fernandes-Henriques C, Li B. 2017. Optogenetic probing of ventral pallidal neurons in motivated behaviors. In *Society for Neuroscience Annual Meeting*. Washington, DC, poster presentation.

Giovanniello J, Ahrens S, Li B. 2017. Disrupted neural circuits in a mouse model of ASD. In *Society for Neuroscience Annual Meeting*. Washington, DC, oral presentation.

Stephenson-Jones M, Bravo-Rivera CC, Fernandes-Henriques C, Li B. 2017. Genetically distinct ventral pallidal neurons encode the motivation for reward approach and punishment avoidance. In *Society for Neuroscience Annual Meeting*. Washington, DC, poster presentation.

Yu K, Ahrens S, Zhang X, Schiff H, Ramakrishnan C, Fenno L, Deisseroth K, Zhao F, Luo M-H, Gong L, He M, Zhou P, Paninski L, Li B. 2017. Central amygdala controls learning in the lateral amygdala. *Nature Neurosci* **20:** 1680–1685.

THE STUDY OF INTELLIGENT MACHINES

P.P. Mitra D. Ferrante, A. Field-Pollatou, S. Georgiou, B. Huo, C. Lee, X. Li, L. Lodato, I. Mallardi, J. O'Rourke, S. Savoia, A. Tolpygo, F. Xu

The goal of our laboratory is to discover and study the principles of intelligence underlying real and artificial brains. This goal is reflected in an integrative research program, including experimental, computational, and theoretical components. We are also pursuing artificial intelligence (AI) applications in automating neuroscientific data analysis and clinical diagnoses.

Dr. Mitra initiated the idea of mesoscale connectivity mapping across vertebrate brains using systematic tracer injections on a grid spanning entire brains. The experimental efforts at CSHL and a collaborative laboratory at RIKEN Brain Science Institute (part of the Japanese Brain Initiative) are directed toward mapping mesoscale connectivity in the mouse and marmoset. A recently started experimental collaboration with MIT/Broad is directed toward spatially mapped single-nucleus RNA sequencing to facilitate a mouse brain cell-type census.

Because mature data sets now exist for both mouse and marmoset, computational efforts to analyze this data are an increasing component of the work. Starting in fall 2017, the laboratory is also providing computational and infrastructural support in the form of a data core and an online data portal to a consortium of researchers based at CSHL that is led by Dr. Josh Huang and Dr. Pavel Osten as part of the U.S. Brain Initiative Cell Census Network. This augments the Mouse Brain Architecture (MBA) Project with cell type–specific data. A new computational area of research for the laboratory in collaboration with physicians at Northwell Health involves computational pathology with a goal of automating anatomic pathology (AP) diagnoses. This research is related to and builds on machine vision techniques developed for analyzing neuroanatomical images.

In addition to the experimental and computational efforts in the laboratory, Dr. Mitra is performing theoretical and mathematical work in machine learning. A focus of this research is to understand the success of modern machine learning methods (deep networks), and involves the application of theoretical tools from statistical many-body physics.

Dr. Mitra is a distinguished (visiting) professor at the Indian Institute of Technology (IIT), Madras, where he initiated a Center for Computational Brain Research and founded a for-credit course entitled "Machine Intelligence and Brain Research." A team of computational scientists at IIT Madras are engaged in assisting the Mitra laboratory in computational and informatics tasks. Dr. Mitra is also a Senior Visiting Researcher at the RIKEN Brain Science Institute in Tokyo, Japan.

Mesoscale Circuit Mapping: MBA Project

The MBA project, initially funded by the National Institutes of Health (NIH) (Challenge and Transformative Research Grants) and subsequently supported by the Mathers Charitable Foundation, aims to develop a brain-wide connectivity map of the adult (8-week-old) male C57BL/6 mouse brain at the "mesoscopic scale." Dr. Mitra originally proposed such a mapping strategy and defined the "mesoscale" as the transitional length scale from a "microscopic" scale at which individual variation is prominent to a "macroscopic" scale at which one can see species-typical patterns. The exact mesoscopic length scale differs with brain region, approximately hundreds of microns to ~1 mm. The mesoscale roughly corresponds to the scale of interest in classical neuroanatomical reference atlases. Armed with the tracer injection data set on a brain-wide grid, the goal is to understand fundamental organizational principles of the whole mouse brain, and more generally of the mammalian brain, and bring about a new conceptual synthesis with long-term impact on neuroscientific understanding.

Experimental Summary

The experimental goal has been to obtain complete injection coverage using one anterograde tracer (adeno-associated virus or AAV) and one retrograde tracer (cholera toxin subunit B or CTB). In 2017, we added a new tracer, an AAV transporting in the retrograde direction (rAAV2-retro, supplied by Dr. L. Looger, HHMI Janelia Research) that permits robust retrograde projections with efficiency comparable to the classical CTB tracer.

Data Analysis Summary

In 2017, we commenced on the analysis of the completed AAV anterograde data set. Repeat injections will be conducted if necessary to increase coverage. The analytical work can be divided into four parts: (a) an improved algorithm for brain to atlas registration; (b) work on the online data portal, particularly improved user interfaces to navigate the large collection of injections from the project; (c) machine vision methods to detect axonal processes in the brain; and (d) topological skeletonization of the tracer injections to summarize the whole-brain projection data in a manner more biologically relevant than simple connectivity matrices.

Registration Pipeline

This work was done in collaboration with M. Miller, B. Lee, and D. Tward (Johns Hopkins University).

An important step is registering the individual brains to a common atlas. In a continuing collaboration with Michael Miller at JHU using techniques from differential geometry (large deviation diffeomorphic metric mapping), we have recently made a significant improvement in our registration pipeline. The previous approach was to first assemble the 2D brain image sections into a 3D stack using section-to-section registration, followed by registration of the 3D stack to the reference atlas. This approach worked, but we noticed that the resulting deformations of space were large. In the new approach, we are using a combined method in which the 2D section registration occurs in tandem with registration with the 3D atlas mapping step, so that the atlas can inform the 2D section-to-section registration as well. This has reduced the deformation of brain space.

Topological Skeletonization of Tracer Injections

This work was done in collaboration with Y. Wang and D. Wang (Ohio State University).

Each tracer injection labels thousands of individual neurons, each one of which is a tree; it stands to reason, therefore, that we could summarize the tracer injection using a tree-like skeleton. This work is being performed collaboratively with Dr. Yusu Wang from OSU using a method from computational topology called discrete Morse theory.

Mesoscale Circuit Mapping in the Marmoset

This work was done in collaboration with Y. Takahashi, M. Lin, J. Nagashima, M. Hanada, and B. Huo (Laboratory at the RIKEN Brain Science Institute).

In 2014, Dr. Mitra was invited to be a part of the Japanese Brain/MINDS project (the Japanese Brain Initiative) and to replicate the MBA Project pipeline at the RIKEN Brain Science Institute in Tokyo to perform a similar approach in the marmoset monkey. The common marmoset (*Callithrix jacchus*) is an emerging primate model system of widespread interest seeing rapid growth compared with the macaque, because of faster generation times (allowing more facile genetic manipulations), a smaller brain with fewer cortical folds allowing for easier access to the cortical surface, and behavioral traits such as a rich social vocal communication repertoire. In 2017, this experimental pipeline was in full operation—to date, 166 injections of retrograde and anterograde tracer substances have been placed in different brain locations, and about 100 of these injections digitized to produce a rich data set that is currently being subjected to computational analysis using techniques developed in the mouse. See Figure 1 for an example brain showing tracer projections.

Whole-Human Brain Histology Correlated with MRI

This work was done in collaboration with Dr. L.L. Latour (NIMH).

In a continuing collaboration with Dr. Larry Latour's group at the National Institute of Neurological Disorders and Stroke (NINDS)/NIH, we have taken

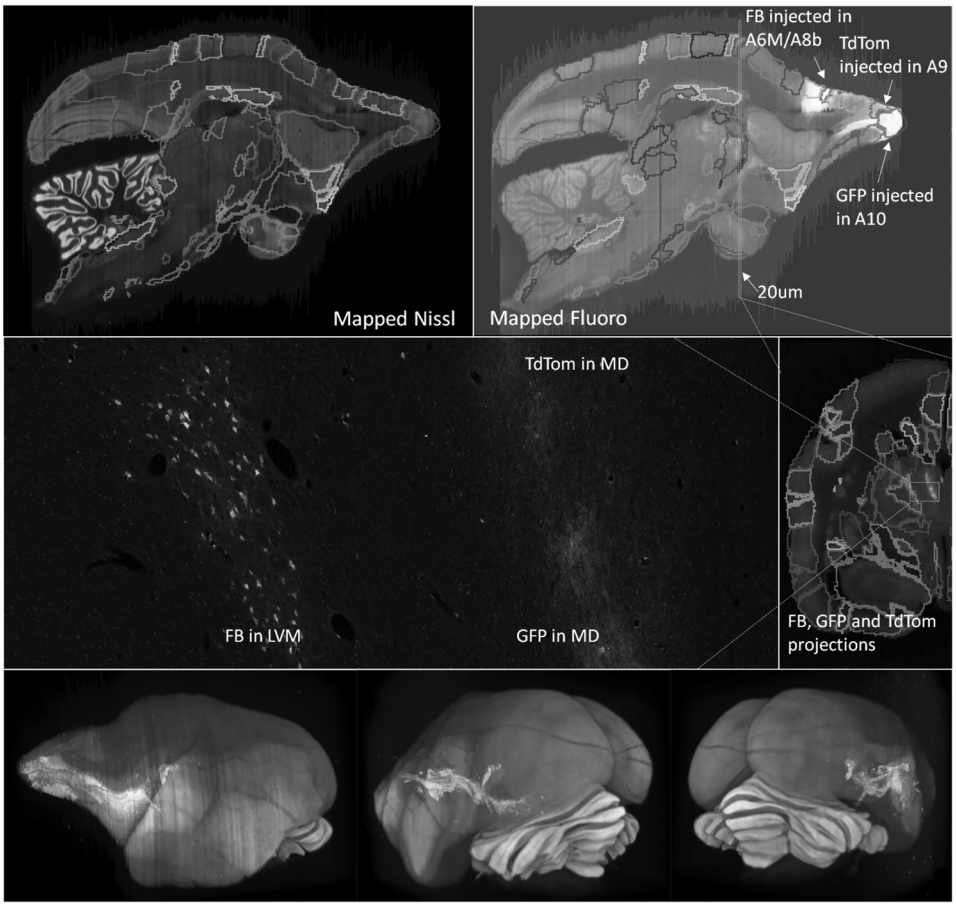


Figure 1. (*Bottom*) 3D reconstructed view of marmoset brain showing three tracer injections (red, green: AAV, anterograde projections; blue: FastBlue, retrograde inputs). (*Top*) Reconstructed 3D volumes together with co-registered atlas segmentation, showing virtual sagittal sections through Nissl-stained and fluorescently imaged coronal sections. (*Middle*) Tracer label in selected projection compartments.

subsequent steps in a larger effort aimed at creating a pipeline for whole human brain digital histology. We have developed unique methods and tools for processing large format specimens up to 6" × 8" in size. Brain tissue was sectioned at 20 μm, and mounted sections were stained with hematoxylin and eosin (H&E) and iron/Prussian blue. Custom large tape, slide holders, staining apparatus, rollers, molds, and protocols were engineered, machined, and tested to improve the speed and consistency of processing inside our whole-body cryostat, the Leica CM3600. As previously reported, this large-format approach allows us to elucidate the morphology of traumatic vascular injury seen in human postmortem tissue and adds microscopic details to findings on the in vivo MRI.

Computational Neuroanatomy, Web Informatics, and Spatial Transcriptomics

This work was done in collaboration with M. Miller (Johns Hopkins University); Y. Wang (Ohio State University); and M. Sivaprakasan, K. Ram, S. Das, IIT Madras; E. Macosko, and A. Regev (Broad Institute/MIT).

The availability of whole-brain histological data in neuroanatomy is relatively new and requires new computational methods. In collaborative work, methods from differential geometry, topology, and machine learning are being incorporated into the analysis of neuroanatomical data, as described in the previous section on mouse mesoscale connectivity mapping. In addition, the presentation, navigation, and annotation of high-resolution histological data sets produces its

own set of challenges that have to be addressed. The effort devoted toward these computational and informatics problems has been increased in the lab.

In 2017, Dr. Mitra started the Data Core component of a U19 award as part of the NIH Brain Initiative Cell Census Network (BICCN) led by Josh Huang (CSHL) and Paola Arlotta (Harvard), also serving a similar role for a U01 award to Pavel Osten. Part of the work in the data core is to centralize project data from different investigators for analysis, web presentation, and for transfer to NIH-designated web archives. Work in the data core can be broken down into (1) data ownership, stewardship, and metadata management, (2) storage and infrastructure management, and (3) portal development and tool enhancements.

In addition to the initially proposed informatics components, the Mitra laboratory is now participating in a web-lab component of the Huang U19, performing spatially localized mapping of single-nucleus RNA sequencing-based cell-type analysis in the mouse brain. This involves an integrated workflow involving the MBA pipeline and Evan Macosko's laboratory at the Broad Institute at MIT. Fresh-frozen mouse brain sections are serially sectioned using tape transfer, and small punches of tissue removed from a few sections for snRNAseq processing. These tissue punches are shipped to the Broad Institute, where they are processed. The remaining tissue sections are processed through the MBA pipeline using the Nissl stain, permitting reassembly into digital 3D histological volumes that can be mapped into a reference atlas. The punches from the tissue samples obtained to perform snRNAseq can then be placed precisely into the reference atlas to obtain a spatially localized analysis of the single-nucleus transcriptomes.

Computational Pathology and Diagnostic Automation

This work was done in collaboration with A. Kepecs (CSHL); J. Crawford, M. Conforti, and D. Savant (Northwell); B. Gallas and Q. Gong (FDA); and C. Abbey (UCSB).

With exponential growth in AI/machine learning, there is significant interest in computer-aided diagnoses (CADs). In the context of AP, diagnoses are based on visual interpretation of microscopic images of H&E stained tissue samples. A new project in the lab aims to make progress in this direction by studying how anatomic pathologists perform diagnostic judgments by monitoring their behavior while they perform a diagnostic task. The goal is to develop machine learning algorithms to automate cancer diagnoses based on microscopic images of tissue biopsies. Our main innovation is to use visual neuroscience techniques to characterize expert human performance in histopathology, combined with advances in machine learning. Quantification of human performance, and the generated training data, will be used to benchmark and improve the diagnostic algorithms.

On the machine learning/AI side, this is a natural direction from the analysis of neurohistological images in the MBA project. This collaborative effort includes contributions from human psychophysics (Adam Kepecs), clinical AP (James Crawford), and the FDA (Brandon Gallas). Preliminary work has been commenced in the project using an apparatus (eeDAP) developed at the FDA, which permits the tracking of the visual field of a pathologist performing a diagnostic judgment while observing and manipulating a tissue slide under a microscope. The eeDAP apparatus permits correlating the physical field of view with the corresponding portion of the digitized microscopic image. Task paradigms have been designed keeping the clinical workflow in mind, and should permit acquisition of domain-appropriate behavioral data from an expert group of anatomic pathologists.

PUBLICATIONS

Chen Y, McElvain L, Tolpygo AS, Ferrante DD, Karten H, Mitra PP, Kleinfeld D, Freund Y. 2017. The active atlas: Combining 3D anatomical models with texture detectors. medical image computing and computer assisted intervention. In *MICCAI 2017. Lecture Notes in Computer Science*, Vol. 10433. Springer, Cham.

Kim Y, Yang GR, Pradhan K, Venkataraju K, Bota M, García M, Luis C, Fitzgerald G, Ram K, He, et al. 2017. Brain-wide maps reveal stereotyped cell-type-based cortical architecture and subcortical sexual dimorphism. *Cell* **171:** 456–469.

Lee B, Tward DJ, Ferrante DD, Ram K, Mitra PP, Miller MI. 2017. Computational anatomy methods for registration of high resolution mouse brain histology using multichannel LDDMM. In *Abstracts: Society for Neuroscience 2017*, Washington, D.C.

Li Y, Wang D, Ascoli GA, Mitra PP, Wang Y. 2017. Metrics for comparing neuronal tree shapes based on persistent homology. *PLoS ONE* **12:** e0182184.

Rosa MG, Majka P, Takahashi Y, Worthy KH, Lin MK, Wolkowicz IH, Tolpygo AS, Bai S, Huo B, Chan JM, et al. 2017. Monosynaptic connections between auditory and visual areas in the marmoset monkey. In *Abstracts: Society for Neuroscience 2017*, Washington, D.C.

Tolpygo AS, Lodato Z, Griffin A, Lodato L, Latour L, Mitra PP. 2017. Tape-transfer enabled 3D digital histology for whole human brains with MRI co-registration. In *Abstracts: Society for Neuroscience 2017*, Washington, D.C.

Wang S, Li X, Wang Y, Mitra PP. 2017. Summarization of a brain-wide data set of anterograde tracer injections in mouse using topological skeletonization. In *Abstracts: Society for Neuroscience 2017*, Washington, D.C.

MAPPING THE BRAIN AT CELLULAR AND SUBCELLULAR RESOLUTION

P. Osten V. Bakshi M. Bota P. Einhorn J. Mizrachi R. Munoz Castaneda A. Narasimhan R. Palaniswamy X. Qi E. Szelenyi J. Taranda K.U. Venkataraju

Mapping the Brain at Cellular and Subcellular Resolution

How are brain circuits built to achieve the vast array of mammalian behaviors, both innate and learned, and emotions? How can we distinguish normal individual-to-individual variation from abnormal brain features that may underlie the risk for neurodevelopmental and psychiatric disorders in the human brain?

A long-standing barrier to addressing these fundamental questions has been the difficulty in surveying the entire brain at cellular or subcellular resolution and the lack of tools to quantitatively compare cell distributions in the brains of different animals. Over the last decade, work in our laboratory focused largely on the development of high-throughput and -resolution methods for standardized, automated, and quantitative analyses of whole-brain anatomy and function. Today, although we continue to pursue further method development, for example, to achieve 3D super-resolution imaging (see Research Program 3, below), most of our current work focuses on applying our methods to gain detailed understanding of brain structure under normal conditions, as well as in disturbed states in genetic mouse models of neurodevelopmental disorders.

Research Program 1. Cell Type–Based Understanding of the Mouse Brain

Brain cell type anatomy informs all topics in the neurosciences: The distribution and ratios of cell types and their wiring into neuronal circuits underlie the vast diversity of mammalian behaviors. Using our automated pipeline of imaging (serial two-photon tomography, STPT) and computational methods for unbiased cell type atlasing in the mouse brain, we have recently published the first quantitative brain-wide maps for seven inhibitory and modulatory cell types proposed to play key roles in essentially all forms of brain processing and cognitive functions (Kim et al. 2017). These data sets offer an unprecedented wealth of quantitative information about the brain-wide distribution of these important neuronal cell types, representing a highly unique resource for the neuroscience community that is accessible via our webportal http://mouse.brainarchitecture.org/ost/ built in collaboration with Partha Mitra's lab at CSHL. Our own analysis of these data already revealed two highly unexpected findings. First, we uncovered a novel hierarchical organization in the isocortex based on varying ratios of two major inhibitory cell types, the parvalbumin (PV+)- and somatostatin (SST+)-positive interneurons, showing for the first time that different cortical areas contain distinct local neuronal circuits. Second, we identified 11 subcortical areas that differ in cell type composition in the male and female brains, 10 of which contained more cells (SST+ or VIP+) in females. This contrasts with previous studies that identify increased total neurons in males in sex dimorphic areas, highlighting the power of unbiased anatomic studies to discover surprising neuronal circuit features.

Based on the success of this work, we were awarded new BRAIN Initiative funding to build a Mouse Brain Atlas of Cell Type Distributions and Morphologies for more than 100 neuronal and glial cell types. As part of this effort, Arun Narasimhan has developed a new microscopy for fast imaging of cleared brains at high resolution, named oblique light sheet tomography (OLST), that combines light-sheet fluorescence microscopy (LSFM) and vibratome sectioning to allow optimal image resolution throughout the entire sample (Narasimhan et al. 2017; this work is a collaboration with Florin Albeanu, CSHL). Combining OLST and sparse cell morphology labeling by systemic delivery of adeno-associated virus (AAV) (Fig. 1), we plan to map single-cell morphologies for

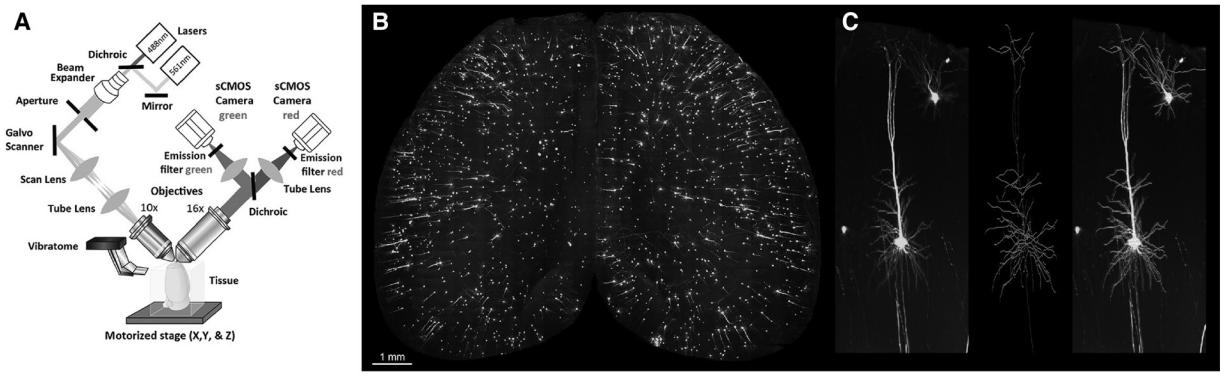


Figure 1. Oblique light sheet tomography (OLST)-based neuronal morphology reconstruction. (*A*) OLST instrument schema. (*B*) Top view of OLST-imaged sparsely labeled pyramidal neuron morphologies in a mouse brain. (*C*) Labeling of individual neurons and semiautomated tracing of their cell morphologies.

hundreds of thousands of cells in the mouse brain, generating an extraordinary amount of structural cell type–based data in both the male and female brains as a resource for the neuroscience community.

Research Program 2. Volumetric Super-Resolution Imaging in the Mouse Brain

Although the above-described methods can image mouse brains at high cellular-level resolution (e.g., at $0.4 \times 0.4 \times 2.5$ microns for the first-generation OLST and $0.4 \times 0.4 \times 1.0$ microns for the second-generation OLST), we still lack technology capable of visualizing and quantifying synapse distribution at nanometer super-resolution scale across 3D tissue volumes. To address this technological need, Judith Mizrachi and Xiaoli Qi are building a second-generation OLST that will integrate super-resolution and cellular-resolution imaging in the same platform. We expect that this method will enable not only better understanding of the relationships between local and local range synaptic circuits, but will also help to answer many hypotheses about structural changes in mouse models of neurodevelopmental and neurodegenerative disorders.

Research Program 3: The Structure and Function of Neural Circuits Involved in Social Behaviors

A key question in systems neuroscience is to understand how brain-wide patterns of neuronal activity relate to sensory encoding, cognitive processes, and ultimately behaviors. Traditionally, the circuit underlying a behavior has been identified by a piecemeal approach: A specific hypothesis about the involvement of a particular brain region is tested by correlating its neural activity with behavior, and the causality of the relationship is tested by loss-of-function or gain-of-function experiments. To dramatically enhance the throughput of such circuit-mapping studies, we have developed an unbiased brain-wide screening platform based on whole-brain mapping of activity-driven IEG (e.g., *c-fos*) expression (Ragan et al., *Nat Methods 9:* 255 [2012]; Kim et al., *Front Neural Circuits 10:* 3 [2016]; Renier et al., *Cell 165:* 1789 [2016]). Applying this method pipeline in the mouse, we have generated the first complete representations of brain-wide activation evoked by two social stimuli: a brief interaction with either a conspecific male or female (Kim et al., *Front Neural Circuits 10:* 3 [2016]).

Building on this proof-of-principle study, Rodrigo Munoz Castaneda has initiated a collaboration with the lab of Steven Phelps at the University of Texas, Austin, to study social behaviors in prairie voles—a rodent species that has become a major model in social neuroscience of pair bonding. Because neither mice nor rats form adult bonds, the pair-bonding behavior of prairie voles offers a unique opportunity to map circuits that are related to monogamous behavior in other species, including humans. As a first step toward this goal, we have developed a pipeline of computational methods for analysis of prairie vole brain anatomy and function (Fig. 2). As this work progresses, we also plan to add other rodent species

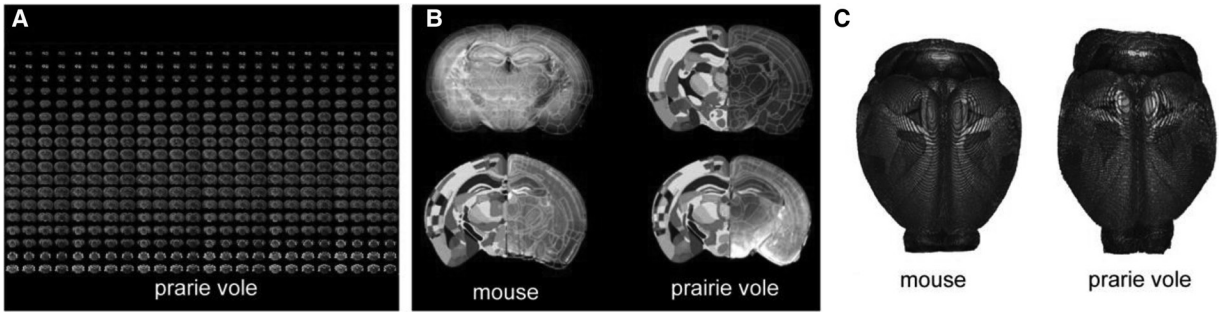


Figure 2. Prairie vole brain-mapping toolbox. (*A*) Overview of 428 coronal sections of serial two-photon tomography (STPT)-imaged prairie vole anatomical brain atlas. (*B*) Region-based sections of the mouse and prairie vole brains imaged by both STPT and the light-sheet fluorescence microscopy (LSFM) and aligned with labels for different anatomical brain areas. (*C*) 3D anatomical representation of mouse and prairie vole brains.

with distinct social behaviors, such as montane voles, which are polygamous and, in collaboration with the lab of Hopi Hoekstra at Harvard, deer mice *Peromyscus polionotus* and *Peromyscus maniculatus*, which are monogamous and polygamous, respectively. By mapping the structure–function relationships in these related rodent species, we hope to obtain a detailed understanding of the brain regions and circuits mediating these markedly distinct and evolutionarily highly significant social behaviors.

Deciphering Mouse Models of Human Neurodevelopmental and Psychiatric Disorders

Identifying vulnerable circuits in mouse genetic models of neurodevelopmental and psychiatric disorders has been a major motivation for the development of the whole-brain mapping and analysis methods in my laboratory. Although most of our work to date has focused on method development and applications to the study of normal brain circuits, we have recently completed two studies focused on genetic conditions linked to abnormal neurodevelopment. In the first study, Eric Szelenyi mapped the distribution of cells with either the maternal or paternal X chromosome active in the female mouse brain, showing a brain-wide bias in X-chromosome inactivation (XCI) causing more cells in the female brain to have the maternal X chromosome active (Fig. 3). Next, Eric tested a condition in which a genetic lesion on the X chromosome—a loss of a copy of the fragile X mental retardation-1 (*FMR1*) gene that causes the neurodevelopmental fragile X syndrome—was inherited either from the mother or the father. Strikingly, these experiments revealed not only that the mutation inherited from the mother caused more pronounced phenotypes than the same condition inherited from the father, but also that in an individual brain, local ratios of maternal versus paternal X active cells played a critical role in determining the exact phenotype, such as increased anxiety versus decreased sociability. These results thus showed that (1) XCI in the female brain is not entirely random and, in fact, shows bias toward more cells with maternal X active, and (2) such XCI distribution can significantly affect clinical phenotypes in X-linked neurodevelopmental disorders, such as the fragile X syndrome.

In a second study focused on neurodevelopmental conditions, Julian Taranda identified a remarkable condition of an incomplete penetrance of structural, epileptiform, and behavioral phenotypes in a mouse model of 16p11.2 deletion, a recurrent copy number variation (CNV) linked to developmental delays, intellectual disability, autism, and childhood seizures. Approximately half of the genetically identical 16p11.2 deletion (16p11.2 del/+) mice showed a number of prominent phenotypes, including spontaneous epileptiform episodes of cortical activity, increased propensity to convulsant-evoked seizures, pronounced volume reductions in cortical areas correlating with increased convulsant-evoked local cortical activity, disrupted sleep, hyperactivity, and increased repetitive behaviors. In contrast, the remaining 16p11.2 del/+ mice showed only moderate brain volume changes and hyperactivity, but otherwise appeared phenotypically normal. These findings are, to our knowledge, the first report of a phenotypic discordance in a neurodevelopmental CNV mouse model, establishing the 16p11.2 del/+ mice as an

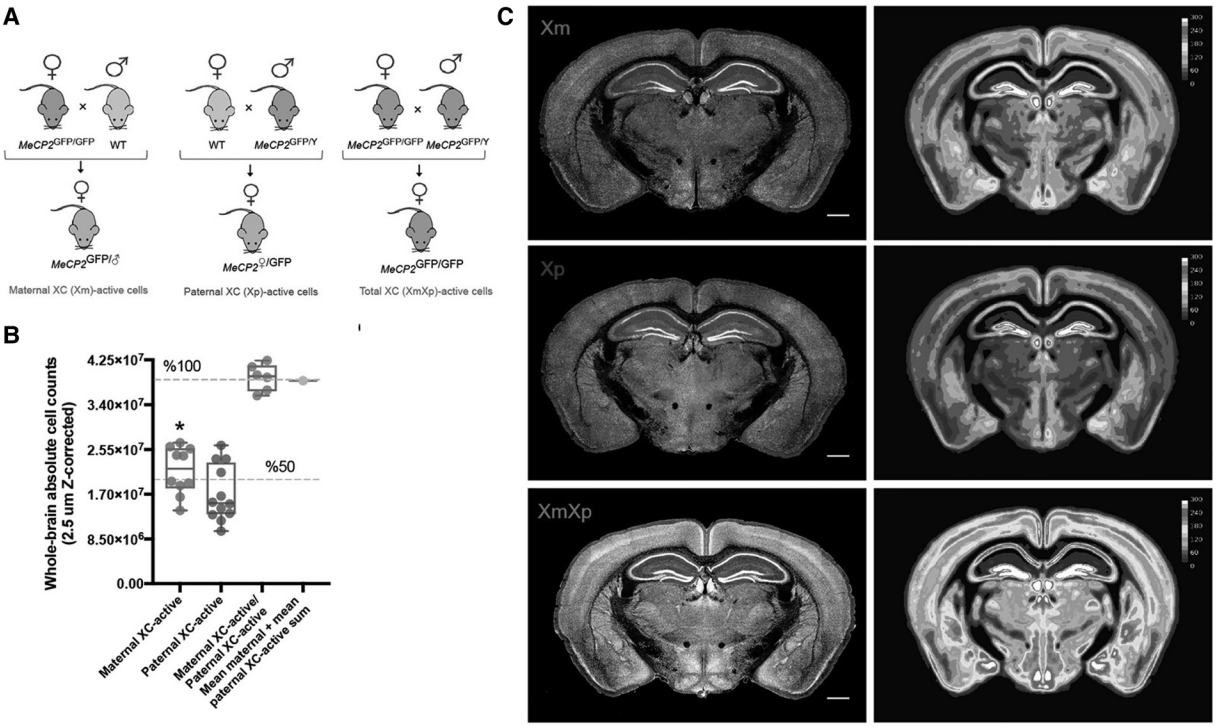


Figure 3. X-chromosome inactivation (XCI). (*A*) Schema of generating mice with maternal, paternal, or both X chromosomes marked by expression of green fluorescent protein (GFP). (*B*) Quantification of the cells with maternal or paternal X activate in the entire brain. (*C*) Examples of X-chromosome-specific cell labeling (*left*) and heat of the densities in the mouse brain.

important tool to study the relationship between genetic and nongenetic factors affecting brain development and neurodevelopmental phenotypes. Finally, we propose that discordance in neurodevelopmental phenotypes is present in other CNV mouse models as well, although it can be frequently missed in traditional two-group comparisons between mutant and wild-type littermate mice.

Summary

Motivated by the gap in knowledge presented by difficulties in quantifying cellular phenotypes across the entire mammalian brain, our lab has pioneered a series of high-throughput and -resolution methods to enable new discoveries relevant to brain circuit functions, gender dimorphism, social behavior, brain evolution, and abnormal neurodevelopment caused by human genetic risk factors.

PUBLICATIONS

Ecker JR, Geschwind DH, Kriegstein AR, Ngai J, Osten P, Polioudakis D, Regev A, Sestan N, Wickersham IR, Zeng H. 2017. The BRAIN Initiative Cell Census Consortium: Lessons learned toward generating a comprehensive brain cell atlas. *Neuron* **96:** 542–557.

Kim Y, Yang GR, Pradhan K, Venkataraju KU, Bota M, Garcia Del Molino LC, Fitzgerald G, Ram K, He M, Levine JM, et al. 2017. Brain-wide maps reveal stereotyped cell-type-based cortical architecture and subcortical sexual dimorphism. *Cell* **171:** 456–469.e422.

Narasimhan A, Umadevi Venkataraju K, Mizrachi J, Albeanu DF, Osten P. 2017. Oblique light-sheet tomography: Fast and high resolution volumetric imaging of mouse brains. bioRxiv doi: 10.1101/132423.

In Press

Turcan S, Makarov V, Taranda J, Wang Y, Fabius AWM, Wu W, Zheng Y, El-Amine N, Haddock S, Nanjangud G, et al. 2018. Mutant-IDH1-dependent chromatin state reprogramming, reversibility, and persistence. *Nat Genet* **50:** 62–72.

NEURAL CIRCUITRY FOR SOCIAL COMMUNICATION

S. Shea B. Cazakoff R. Dvorkin A. Nowlan L. Shen
A. Corona C. Kelahan D. Rupert

The broad goal of my laboratory's research is to understand how the brain detects and interprets sensory stimuli to guide flexible behavior. We are particularly interested in how neural activity and plasticity in olfactory and auditory brain circuits facilitate communication and social behavior. We are revealing neural mechanisms that allow organisms to detect and recognize familiar individuals, gather information about their identity and social status, and select appropriate behaviors. Mice can acquire a surprisingly detailed profile of a social partner from the smells and sounds it emits during their encounter. For example, they can discern the other mouse's sex, genetic identity, reproductive state, levels of distress or sexual interest, or even recently consumed foods. As you might imagine, proper interpretation of these social signals is indispensable for survival and mating success.

It is quite reasonable to ask, "Why do you care about the social life of small rodents?" There are really two reasons. First, we want to identify the fundamental principles that govern how the brain adaptively controls complex behavior. Natural social behaviors are well suited for this purpose because the brain is exquisitely adapted through evolution to resolve and integrate social cues and link them to powerful behavioral responses. Second, we hope to pinpoint and repair neural circuitry defects that impair appropriate use of social information. Difficulty with social perception and cognition are core features of the autism spectrum disorders (ASDs). For example, patients may have trouble perceiving and interpreting communication gestures such as speech, facial expressions, and "body language." This broad feature is also evident in many mice that carry genetic variants identified in human ASD populations. Therefore, if we can ascertain the neural circuit substrates of social behavior in normal mice, we can make and test predictions for how the circuitry is affected in the mouse models. The results are likely to tell us more about the synaptic modifications that occur in human autism.

In recent years, we have focused on understanding the neural activity and circuitry in primary sensory brain areas that support adaptive behavior. We are now moving deeper into the brain to ascertain where the sensory data from those regions are collected and integrated into hormonal and electrical signals that promote appropriate behavioral choices. We are particularly interested in how all stages of neural processing are flexible according to experience and behavioral state.

Auditory Plasticity Sharpens Vocal Perception during Parental Learning

A. Corona, A. Nowlan, L. Shen

Far beyond the limits of our hearing, in the ultrasound range, mice are continuously "speaking" (or vocalizing) to one another in a "language" that we have only just begun to understand. Many distinct vocalization types are produced by males, females, juveniles, and adults in a variety of behavioral situations. We would like to better understand the perceptual significance of these calls to the mice and how they are used to guide behavioral choices. One form of vocalization that is reasonably well understood is the ultrasonic distress vocalization (USV). Young mice, before vision and full mobility, will occasionally become separated from the nest. This is stressful for them and they will, therefore, call out to their mother with a very high-frequency peep. New mothers develop sensitivity to these cries and respond by moving toward their source to retrieve the pup. Moreover, females who have never given birth do not innately show approach responses to pups or their calls; however, they can learn to perform the behavior with experience. We refer to this group as "surrogates." Accurate maternal learning by mothers and surrogates is suspected to involve rewiring (or "plasticity") in the auditory cortex.

One potentially valuable approach for understanding this plasticity and its behavioral consequences is to thoroughly explore the perceptual boundaries that define a vocal category. If a sensitive, high-throughput method for achieving this were available, we could

identify those behavioral boundaries, compare them with the neural boundaries that define selective responses to USVs, and monitor how they altered over time as an animal gains maternal experience. Graduate student Luqun Shen is developing a behavioral assay that he can use to probe the necessary features that define USVs with higher resolution than previously possible. Once he does that, he will be able to assess how those features change, and how responses to those features in the brain change after a female mouse is given maternal experience.

Graduate student Alexandra Nowlan is hoping to visualize auditory cortical plasticity day by day in a mouse learning to care for pups by literally peering through a window into the brain. Together with technician Clancy Kelahan, she has begun implanting glass windows that allow her to use neural imaging techniques to take daily snapshots of the brain's response to pup calls. Now that she has perfected the method, Ally will begin tracking changes in these patterns as a female mouse is growing more proficient in care of the pups. Another aspect of her project is her identification of a novel input to the auditory cortex from a region of the amygdala that processes social odors. Preliminary data Ally has gathered suggests that this pathway actually modulates sound processing according to these odors, and the pathway may be important for maternal learning. Our working model is that this pathway combines olfactory stimuli, such as the smell of pups, with auditory stimuli, such as the vocalizations of pups.

The vast majority of studies of parental behavior in mice have focused entirely on females and maternal behavior. Nevertheless, graduate student Alberto Corona has shown in preliminary data that, under certain narrower conditions, males will show paternal behavior. Alberto is very interested in identifying contextual factors that favor paternal behavior. We are hypothesizing that vocal and olfactory cues emitted by the mother may be very important for encouraging paternal behavior. He is also interested in identifying brain regions in males that overlap with circuits that govern maternal behavior, as well as brain regions that are uniquely activated during paternal behavior. Therefore, Alberto is performing behavioral experiments and whole-brain imaging (in collaboration with P. Osten) from male and female mice to screen for paternal behavior-specific neural activity.

Auditory Plasticity and Parental Behavior Are Impaired in a Mouse Model of Rett Syndrome

D. Rupert [in collaboration with K. Krishnan and B. Lau, CSHL]

Postdoctoral fellows Billy Lau and Keerthi Krishnan led a related collaboration with CSHL professor Dr. Josh Huang. They examined how vocal perception of pup calls, and the accompanying neural plasticity, are affected in mice that are missing one copy of a gene called *MeCP2*. Impairments in the function of this gene are understood to cause the ASD Rett syndrome. Indeed, we find that females that possess only a single copy of *MeCP2* are not able to develop proficiency at gathering pups. In our first paper, published in 2017, we showed that this likely happens because MECP2 (the protein product of the gene) plays a critical role in maintenance and plasticity of the auditory cortex by acting on inhibitory interneurons. In a particularly exciting set of experiments, we were able to repair inhibitory function with genetic and pharmacological manipulations, thereby restoring maternal gathering behavior. This finding suggests that there may be a way to achieve cognitive improvement in humans, even after brain development.

We are now preparing a manuscript reporting results of a second study focused on the specific activity changes in different types of auditory cortical neurons during maternal learning. Based on our recently published findings, we predicted that one of the central participants in auditory cortical plasticity is a network of inhibitory neurons that express a protein called parvalbumin (PV). Notably, deletion of *MeCP2* only in this minority of neurons is sufficient to disrupt pup care. Therefore, Billy made neuronal recordings in awake behaving females of both genotypes that differed in their maternal experience. The data show that when a normal adult female mouse is first exposed to pups, her auditory cortex becomes "disinhibited" (i.e., there is suppression of the inhibitory network). This seems to be attributable specifically to suppression of the PV neurons, but remarkably, this disinhibition appears only during the response to actual calls, not in response to synthetic stimuli. Therefore, maternal behavior-stimulated changes to the auditory network are selective for behaviorally relevant stimuli, while leaving the responses to behaviorally irrelevant stimuli unaffected. In contrast to typical mice, we observed no disinhibition in *MeCP2*-deficient mice. However,

mice that received our genetic restoration of inhibitory function also showed restoration of this disinhibition to calls. The results of this study are consistent with our model that MECP2 regulates plasticity in adults and juveniles through its effects on inhibitory interneurons. Graduate student Deborah Rupert is continuing with behavioral and electrophysiological studies that will genetically dissect the contributions of several different classes of interneurons, as well as excitatory neurons.

Neural Activity Signaling Emotion, Arousal, and Reward during Social Encounters

R. Dvorkin

Organisms are constantly being bombarded by an overwhelming number and variety of stimuli from all of their senses. Therefore, one of the greatest challenges faced by the nervous system is to make sensible and efficient choices about which stimuli to attend and remember. Decades of evidence have established that this calculation is achieved, in large part, with the contribution of a class of neurochemicals that enable neuronal communication and are collectively referred to as "neuromodulators." Neurotransmitters of this type generally do not participate in fast, moment-to-moment communication between neurons. Instead, their levels fluctuate more slowly and modify the function of larger groups of neurons, or circuits. One important function of neuromodulators is to alert the brain to the possibility that something noteworthy is happening in the environment. Similarly, they may also broadcast a signal that the animal encountered a reward. This is likely to motivate the animal to examine the events leading up to that reward so that it may build a model of the actions and stimuli that predict positive outcomes. Of course, the other side of this coin is that the animal also wants to keep track of recent and impending negative events. In that sense, neuromodulators can inform the brain prospectively that something interesting is about to happen or retrospectively that something interesting recently happened. We are specifically interested in a small nucleus in the brainstem called locus coeruleus (LC) that sends neuromodulatory signals throughout the brain by releasing the chemical noradrenaline (NA). Evidence strongly suggests that in artificial, operant learning tasks, the activity of LC neurons both prospectively influences behavior and retrospectively signals recent stimuli of behavioral interest. However, very little is known about how these neurons participate in unstructured social interaction.

The ability to detect and act on rewards and punishments is an indispensable component of adaptive behavior. Rewards often take the form of food, drugs of abuse, or, in the case of humans, currency. However, a variety of sensory social cues can also be rewarding. Extensive evidence establishes that engaging in social interactions, such as maternal behavior or sexual behavior, is powerfully rewarding to rodents, including mice. The desire for access to these activities can be harnessed to condition changes in behavior. Remarkably, early postpartum interactions with pups can even be a more powerful reward than cocaine. The rewarding properties of social interaction are at least partly driven by the rewarding nature of specific sensory cues experienced during a social encounter. Olfactory stimuli from conspecifics show robust reward characteristics in mice. For example, male-soiled bedding, adult male courtship vocalizations, and the ultrasonic cries of their pups are typically strongly attractive to female mice. In the course of natural free behavior, mice are likely to occasionally encounter threatening or aversive stimuli (e.g., predator odors) as well. Neurons in LC are known to be responsive to both positive and negative events that evoke strong emotion and/or arousal.

Former postdoctoral fellow Dennis Eckmeier successfully developed reliable methods for making stable recordings of individual neurons during actual social encounters between mice, including courtship and parental interactions. This work is now being continued by current postdoctoral fellow Roman Dvorkin. Their specific goal is to monitor the activity of noradrenergic neurons in LC during these encounters. These neurons are likely responsive to rewarding social signals and may modulate coding of sensory data. Understanding the context-dependent activity patterns in LC is, therefore, critical to developing models for how this structure affects behavior. In the past year, Roman has succeeded in chronically observing LC neural activity, day after day, as the mouse has repeated social interactions using two independent methods. He has made electrical recordings of individual LC neurons and also used optical techniques to image patterns of firing among the genetically defined noradrenergic cell class. Both methods clearly reveal transient bursts of

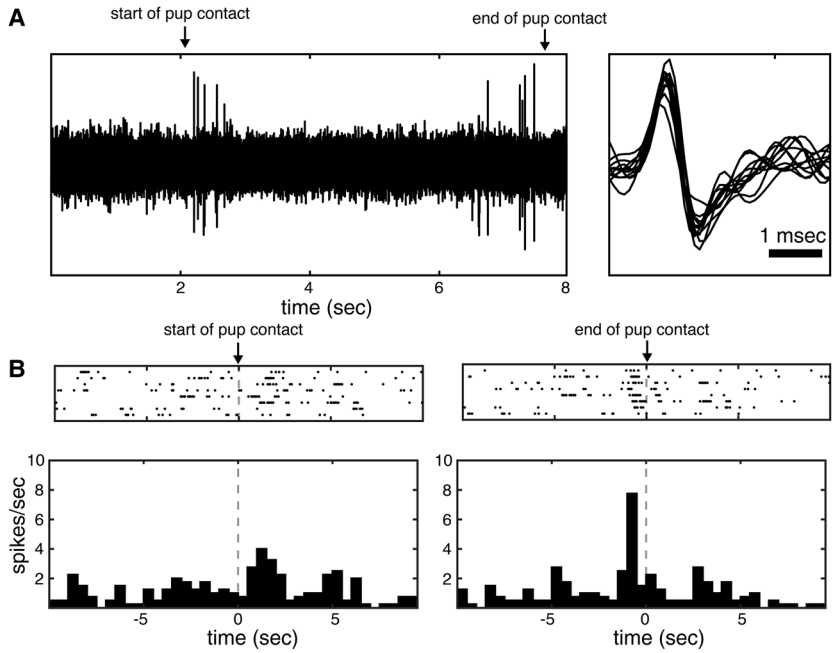


Figure 1. Activity of a single neuron in the locus coeruleus (LC) of a freely behaving female mouse during pup retrieval. (*A*) Raw voltage traces from a presumed LC neuron recorded from a female mouse while retrieving a pup. (*Left*) Approximately 8 sec of recording during a single pup retrieval event. The arrowheads denote the times when the pup was picked up and when it was dropped in the nest. (*Right*) Plot of all isolated spikes from the *left* panel overlaid to show the consistent shape and amplitude. (*B*) This LC neuron consistently bursts after the initiation and before the termination of each retrieval event. (*Left*) The raster plot and peristimulus time histogram depict the activity of the presumed LC neuron, aligned to the time the pup was picked up during each of nine separate retrieval events. (*Right*) The raster plot and peristimulus time histogram depict the activity of the presumed LC neuron, aligned to the time the pup was dropped at the nest during each of nine separate retrieval events.

activity associated with positive social events, such as contact with offspring (Fig. 1), or aversive events, such as exposure to a predator odor.

Dynamics of Neuronal Inhibition during Olfactory Learning

B. Cazakoff [in collaboration with D. Kepple and A. Koulakov, CSHL]

Behavioral state and previous experience can dramatically change sensory responses. Rather than a one-way street, sensory perception is probably better viewed as an ongoing and dynamic interaction between the brain and its environment. This interaction can allow sensory activity to be optimized in accordance with behavioral demands, and can help the brain converge to a more accurate and efficient representation of complex stimuli. For the past several years, we have been addressing the contribution to this flexibility of a crucial group of inhibitory interneurons in the main olfactory bulb (MOB) called granule cells (GCs). Our studies combine extensive collection of experimental data from GCs in awake and anesthetized mice and theoretical modeling of GC dynamics during learning in collaboration with the research group led by our CSHL colleague Dr. Alexei Koulakov.

Detection and interpretation of olfactory stimuli in the brain begin at the MOB. Sensory receptor neurons in the nose relay information about odors to projection neurons that constitute the output of the MOB to deeper brain structures such as the olfactory cortex. Despite the rather peripheral location of the MOB, one synapse from the primary sensory neurons, the bulb is heavily targeted by signals that originate deeper in the brain. These "feedback" inputs are important for maintaining and updating sensory responses in a changing environment. Among the neurons in the MOB, the inhibitory GCs are certainly the most numerous, arguably the most important, and yet somehow they are the most mysterious. GCs are the primary target of feedback

connections, making them a conduit for the olfactory bulb to integrate odor information with signals originating deep in the brain. Furthermore, functional studies suggest that they are necessary for olfactory discrimination and learning. Thus, GCs are a crucial link between the projection neurons that carry odor information out of the MOB and central feedback that carry signals necessary for learning. We have two studies ongoing in the lab that are focused on GCs.

First, we developed a task in head-fixed mice that allows us to monitor the neuronal activity of GCs as the mouse rapidly adjusts its behavior in response to changes in odor meaning. We are specifically interested in how the mouse's ability to solve this task involves the GCs. Brittany Cazakoff made recordings of GCs in mice that are actively learning new associations of odor stimuli with reward (water) and punishment (bitter taste). She developed a behavioral paradigm in which the mouse learns that two new odors each signal either impending reward or the bitter tastant, respectively. They are also asked to learn that the odor associations have switched. The mice show that they have learned by licking when reward is signaled and withholding when the odor portends bitter taste. This learning develops in a time frame that allows us to monitor activity in individual GCs throughout. These experiments reveal several very intriguing preliminary findings.

1. Some of these mice successfully learned the task, whereas some did not. Interestingly, we see stronger responses to odors in GCs from mice that had successfully been trained as compared with mice that did not learn the task.
2. Many cells change their response to an odor when it begins to predict a new outcome.
3. We found that responses often fluctuate over time in the trials following this switch. These fluctuations were not random but rather were related to the animal's fluctuations in task accuracy.

Second, we have worked in collaboration with the Koulakov laboratory to develop a model of GCs and their synaptic interactions with the mitral cells (MCs) that convey odor information to deeper brain structures. We speculate that the interaction between these circuits enables the MOB to implement an algorithm for "compressed sensing" that endows the structure with an ability to compress and decompress its representations of many-dimensional stimuli. Compressing sparse multidimensional stimuli would allow the brain to most efficiently represent a vast and diverse chemical space with a limited number of receptors. This compressed sensing scheme has the added benefit that the MOB is capable of learning the identity of components in a mixture without ever experiencing the components in isolation. The GCs are absolutely central to this computation, and our model makes unexpected predictions about their activity during odor learning. We, therefore, compared these predictions with experimental data from GCs collected during a learning paradigm. We found that the empirical observations matched closely with the patterns predicted by our model. Thus, they corroborate our speculation that the MOB refines an implementation of a compressed sensing mechanism through a learning-dependent process.

PUBLICATION

Krishnan K, Lau BYB, Ewall GE, Huang ZJ, Shea SD. 2017. MeCP2 regulates cortical plasticity underlying a learned behaviour in adult female mice. *Nat Commun* **18:** 14077.

SEX DIFFERENCES IN THE BRAIN: FROM GENES TO BEHAVIOR

J. Tollkuhn R. Bronstein M. Wu

Our lab seeks to understand the mechanisms that shape and regulate sex differences in the brain. Females and males differ in many behaviors and are differentially affected by mental health disorders, but the distinct developmental trajectories that give rise to these sex differences remain poorly understood. Much of our knowledge about the cellular and molecular differences between the sexes in the mammalian brain has been obtained through studies of the hormonal regulation of the differentiation and function of neural circuits underlying innate, sex-typical behaviors and physiology in rodents. Paradoxically, estrogen is required to both feminize and masculinize the brain. Males undergo a transient perinatal testosterone surge and this circulating testosterone is converted to estradiol (the most abundant endogenous estrogen) locally in the brain. Treating females with estradiol at birth irreversibly masculinizes both adult behaviors and gene expression patterns, suggesting that perinatal estrogen directs gene regulatory events that organize persistent sex differences in the brain. These long-lasting effects suggest that neonatal estrogen organizes the developing brain through an epigenetic mechanism whereby the transient signal at birth irreversibly modifies the chromatin state of genes. The nature of this mechanism, as well as the identity of the genes regulated by gonadal hormones in the brain, is the focus of our current research program.

Genetic Dissection of a Sexually Dimorphic Neural Circuit

As a first step in understanding the specificity of estrogen signaling in brain sexual differentiation, we performed the first neural-specific deletion of ERα (Esr1) in males. Constitutive deletion of ERα abolishes fertility, alters hypothalamic-pituitary-gonadal (HPG) axis signaling, and impairs sexual and territorial behaviors. These innate behaviors are primarily regulated by hypothalamic regions that receive pheromonal chemosensory information from the accessory olfactory pathway. ERα is expressed in many of these areas, including the medial amygdala (MeA), the principal nucleus of the bed nucleus of the stria terminalis (BNST), the medial preoptic area (MPOA), and the ventrolateral region of the ventromedial hypothalamus (VMHvl). ERα-expressing neurons of the VMHvl and posterior ventral medial amygdala (MePV) are largely glutamatergic, whereas ERα cells in the MPOA, BNST, and posterior dorsal medial amygdala (MePD) are predominantly GABAergic. These regions are organized into a sexually dimorphic circuit whereby chemosensory information from conspecifics is sent from the accessory olfactory bulb to the MeA and BNST, which are interconnected, and project to hypothalamic regions such as the preoptic area (POA) and VMHvl, and additional output areas.

To dissect the contribution of ERα in masculinization of this innate circuit, we deleted ERα/Esr1 in either excitatory glutamatergic neurons or inhibitory GABAergic neurons and assessed alterations in behavior and gene expression in mutant males. Surprisingly, we find that males lacking ERα in excitatory neurons display wild-type levels of male mating, aggression, and territory marking, indicating that although ERα is abundantly expressed in excitatory neurons, including those in the MePV and VMHvl, it is not required in these neurons for the masculinization of behavior. In contrast, deletion of ERα in GABAergic inhibitory neurons alters mating and aggression behavior and abolishes male-typical territorial marking. Twenty-five percent of mutant males attack females, a behavior never displayed in wild-type animals, suggesting that these mutants have difficulty discriminating between the sexes. Furthermore, males lacking ERα in inhibitory, but not excitatory, neurons have dysmasculinized gene expression of two key sexually dimorphic genes. Androgen receptor (AR) is expressed more highly in males compared with females, whereas estrogen receptor beta (ERβ/Esr2) is more abundant in females. Mutant males show decreased AR and increased ERβ. Together, these results suggest that ERα masculinizes the brain by organizing inhibitory inputs from BNST

and MeA onto glutamatergic neurons that drive behavioral output. Thus, it is the information processing and integration areas that possess the greatest sex differences, rather than the hypothalamic areas that have a similar capacity to regulate the display of sex-typical behaviors.

Identification of Sex-Specific Gene Programs in the Brain

To understand how neonatal estrogen directs sexual differentiation of the BNST and MeA, we have performed transcriptomic profiling of these regions in male and female pups. To circumvent the heterogeneity of the mammalian brain, we are performing our RNA-sequencing (RNA-seq) experiments specifically in ERα+ neurons. This specificity is achieved through use of the RiboTag mouse allele, which permits Cre-defined tagging of ribosomes, followed by purification of the affiliated translating mRNAs. We have identified novel genes with sex-biased expression in these regions and noted several genes implicated in autism spectrum disorder (ASD). We therefore collaborated with Ivan Iossifov to assess the significance of this observation. ASD is diagnosed in boys four times as often as in girls, and the origin of this sex bias is a topic of intense investigation. We find a striking enrichment of ASD candidate genes with known "likely gene-disrupting" (LGD) mutations in our female-biased but not male-biased gene lists (Fig. 1). This result suggests that increased expression of specific genes in females may be protective against the effect of a heterozygous LGD mutation, consistent with previous reports that females diagnosed with ASD carry a greater mutational burden than do ASD males. Our finding shows that our strategy of characterizing hormone-responsive neuronal populations is effective at revealing gene programs that underlie sex differences in disease.

Sex Differences in the Epigenome

Our lab is particularly interested in identifying the *cis* regulatory elements that impart sex specificity to gene expression. To gain access to chromatin from our limited populations of ERα neurons, we are using a protocol termed isolation of nuclei tagged in specific cell types (INTACT). This approach uses a Cre-inducible nuclear envelope tag to permit purification of chromatin from genetically defined cell types. The chromatin accessibility landscape of a cell both reflects its developmental history and defines its current transcriptomic capacity. We are using the assay for transposase-accessible chromatin (ATAC-seq) to identify accessible chromatin regions in ERα+ neurons from BNST and MeA. As with our RNA-seq, we are performing our experiments in male and female pups, aged P14. In addition, we are asking if the two sexes have a differential capacity to respond to the same estradiol stimulus in adulthood. We gonadectomize adult animals and, after three weeks, administer a single dose of estradiol or vehicle. We find that adult animals possess more sex differences in chromatin accessibility compared with pups, and many of these differences are present regardless of hormonal status. Four-way statistical analysis with DESeq2 identifies regions that are opened by estradiol specifically in one sex or the other. Recent ChIP-seq studies in diverse cell types and tissues have revealed unexpected strategies for gene regulation by ERα, including recruitment to DNA via other transcription factors, and widespread DNA occupancy in the absence of ligand. We are currently performing motif analyses

			female 212 genes			male 124 genes		
		N	O	E	p-value	O	E	p-value
autism	LGDs	606	30	10.48	1.30E-06	3	1.15	0.2201
	mis	3069	114	76.2	6.40E-05	10	8.91	0.8009
	syn	1198	39	33.24	0.359	3	4.06	0.845
intellectual disability	LGDs	34	0	0.4	1.000	0	0.04	1
	mis	104	6	2.9	0.149	0	0.34	1
	syn	17	0	1.27	0.564	0	0.15	1
schizophrenia	LGDs	95	2	2.55	1.000	1	0.28	0.4884
	mis	580	20	18.54	0.795	1	2.17	0.7255
	syn	218	8	8.09	1.000	0	0.99	0.7456
unaffected	LGDs	224	8	6.09	0.538	1	0.67	0.9754
	mis	1473	43	44.28	0.926	5	5.18	1
	syn	625	23	19.32	0.458	1	2.36	0.6357

Figure 1. Female-biased genes in pup medial amygdala (MeA) estrogen receptor alpha (ERα) neurons are enriched for genes with de novo mutations in autism spectrum disorder (ASD) patients. Table of known de novo mutations (number = *N*) identified in familial sequencing studies, grouped by likely gene-disrupting (LGD), missense (mis), and synonymous (syn) mutations, with expected (E) and observed (O) numbers of mutations, given our sex-biased lists. The 212 female-biased, but not the 124 male-biased, genes expressed in P14 MeA ERα neurons are highly enriched (shadow box) for genes known to have mutations in patients with autism, but not mutations found in other mental disorders or in unaffected siblings.

of our differential peaks to identify the transcription factors that cooperate with ERα in neurons. Furthermore, we are cloning out estradiol-opened peaks to validate enhancer activity in reporter assays, thereby defining novel hormone-responsive genomic regions in neurons. Collectively, these experiments identify enhancers that direct sex differences in gene expression, and implicate specific transcription factors in their regulation.

PUBLICATIONS

Wu MV, Tollkuhn J. 2017. Estrogen receptor α is required in GABAergic, but not glutamatergic, neurons to masculinize behavior. *Horm Behav* **95:** 3–12.

In Press

Manoli DS, Tollkuhn J. 2018. Gene regulatory mechanisms underlying sex differences in brain development and psychiatric disease. *Ann NY Acad Sci* doi: 10.1111/nyas.13564.

CIRCUITRY UNDERLYING CORTICAL PROCESSING AND DECISION–MAKING

A. Zador B. Burbach, F. Carnevale, X. Chen, A. Funamizu, S. Ghosh, H. Gilbert, G. Henry, L. Huang, J. Kebschull, C. Krasniak, S. Lu, F. Marbach, A. Reid, C. Stoneking, Y. Sun, V. Vasily, A. Vaughan, H. Zhan, A. Zhang

Our laboratory is interested in how neural circuits underlie cognition, with a focus on the cortical circuits underlying auditory processing and decision-making. To address these questions, we use a combination of molecular, electrophysiological, imaging, and behavioral approaches. In addition, we are developing a method that will allow us to use high-throughput DNA sequencing to determine the wiring diagram of the brain at single-neuron resolution rapidly, efficiently, and at low cost.

High-Throughput Interrogation of Cortical Activities and Projections at Cellular Resolution

A. Vaughan, C. Stoneking

Corticocortical projection neurons in the mouse primary visual cortex showed activities tailored to their projection targets. It is not known, however, if such target-specific functionality can be generalized to other types of cortical projection neurons or cortical neurons in other cortical areas. To answer this question, we are combining in vivo two-photon calcium imaging with in situ Barseq and MAPseq to examine both the activities and the projection targets of cortical neurons. In a proof-of-principle experiment, we showed that single neurons can be accurately identified both in the two-photon images and the brain slices used for in situ sequencing. We are currently developing a full experimental pipeline with auditory cortex as the area of interest.

Stimulus and Outcome Expectations in Perceptual Decision-Making

A. Funamizu, F. Marbach

In perceptual decision-making, humans and animals use the prior knowledge of stimulus probability and reward outcome to optimize their behavior. Although a series of studies have shown the neural substrate of outcome expectation in the corticobasal ganglia circuit, the neural substrate of stimulus expectation is unclear. To dissociate the two neural circuits, we are training head-fixed mice on an auditory frequency discrimination task (based on Marbach and Zador 2017), in which either (i) the stimulus probability or (ii) the reward size for categories A and B trials changes in blocks, and we analyze the choice behavior with a belief-updating reinforcement-learning model.

We find that both the stimulus probability and reward amount biased mice toward choices associated with high-probability stimuli or large reward, respectively. In the stimulus probability task, a belief-updating model, which recursively updated the probability of stimulus category, predicted the biased behavior better than a reinforcement-learning model, which updated the expected outcome of each category. In contrast, in the reward amount task, the reinforcement-learning model predicted the behavior better than the belief-updating model. This suggests that mice updated their stimulus expectation in the stimulus probability task, whereas they updated their outcome expectation in the reward amount task. We expect that our paradigm combined with two-photon calcium imaging will help dissociate the neural circuits for stimulus and outcome expectations in the auditory cortex.

Comparing Roles of Primary Visual and Auditory Cortices in Decision-Making

A. Zhang

Sensory information of different modalities is processed and extracted through a distinct but analogous series of hierarchically organized brain areas. Visual information travels from the retina to the lateral geniculate nucleus of the thalamus and then on to the primary visual cortex (V1). Auditory information

travels similarly from the cochlea through the medial geniculate nucleus of the thalamus to the primary auditory cortex (A1). Feature-encoding principles are fairly analogous at these early stages of processing, and perceptual decision-making work has begun to elucidate the mechanisms by which such features can then be associated with appropriate motor outputs. However, because of varied stimulus complexities, task demands, and animal models, it has been impossible to directly compare the mechanisms that underlie visual-based associations with those that underlie auditory-based associations.

To overcome these challenges, we have developed a novel visual discrimination task that asks animals to judge the dominant spatial location of the stimulus. We believe that this task is directly analogous and thus permits comparison to an existing auditory discrimination task used in the lab that asks animals to judge the dominant frequency of an auditory stimulus. Animals are able to acquire this new task quickly and to high levels of performance. We are currently using this task to investigate the information carried by neurons in V1 in animals performing a simple visual discrimination. Comparing these with representations in A1 of animals performing the analogous auditory discrimination task, we can compare the contributions of primary sensory areas of different modalities within a common perceptual decision-making paradigm.

Spatial Attention in Mouse Visual Cortex

C. Krasniak

Attention is an essential cognitive function that allows us to select which information from the environment we perceive. When attention is impaired or misdirected, it can greatly impair how we interact with the world around us. Decades of electrical recordings in the monkey visual cortex have elucidated some of the key neural features of spatial visual attention in the cortex—including an increased representation of the attended signal and a decrease in the amount of noise. Other work in humans, nonhuman primates, and rodents has implicated acetylcholine release from the basal forebrain in generating the neural and behavioral effects of attention. We are developing a novel visual spatial attention task in mice to probe this relationship. Using this task, we will leverage the genetic tools of mice to dissect the neuronal mechanisms of spatial attention in mice.

Mapping Whole-Brain Corticocortical Projections in Individual Animals at Single-Cell Resolution by Multisource MAPseq

L. Huang

Long-range neuronal projections determine how information flows among various brain areas to generate behavior, but the logic of information flow in the brain remains poorly understood. Disruption of long-range projection is implicated in neuropsychiatric disorders, including autism and schizophrenia. A central barrier to deciphering this logic is methodological. Conventional anatomical methods have low throughput and usually lack single-cell resolution. In typical experiments, the projections from only a single area—in the limit, only a single neuron—and can be determined in one experimental subject. Mapping the connectivity of an experimental model system is thus very expensive and labor intensive.

We have recently published a method, MAPseq (multiplexed analysis of projections by sequencing), with qualitatively higher throughput for mapping long-range projections at single-cell resolution (Kebschull et al., *Neuron 91:* 975 [2016]; Han et al. 2018). In the first-generation MAPseq, a viral library encoding random RNA barcodes is injected into a single brain area, enabling the multiplexed mapping of approximately 1000 neurons in a single experiment. We now go beyond that. We have increased the throughput of MAPseq by almost two more orders of magnitude and extended it to allow mapping multiple source brain areas. By optimizing viral vectors, injections, dissections, molecular experiments, and the bioinformatics pipeline, with multisource MAPseq we have been able to map the projections—at single-neuron resolution—of more than 50,000 neurons spanning the whole cortex in one individual animal within two weeks at a cost of about $10,000. In contrast, conventional methods require years, thousands of animals, and millions of dollars and do not achieve single-neuron resolution. We are currently applying it to study animal models of neuropsychiatric disorders like autism and schizophrenia.

Corticostriatal Plasticity Underlying Learning of Stimulus–Motor Association in an Auditory Discrimination Task

S. Ghosh

Animals use complex sensory cues from their environment to make a variety of decisions in their lives. Such behavior requires bringing together of sensory discrimination, decision-making, and appropriate motor actions. An auditory discrimination task called the "tonecloud" task was established in our lab to understand the brain circuits involved in such decision-making and how these circuits evolve during learning of sensorimotor associations. Previous studies from the lab have shown that the connections between auditory cortex and the auditory striatum in rats are instrumental for an animal to perform this task (Znamenskiy and Zador, *Nature 497:* 482 [2013]). Moreover, learning of this discriminatory task results in formation of a memory of the learned association in this circuit (Xiong et al., *Nature 521:* 348 [2015]). Nevertheless, the cellular targets of this learning process remain largely unexplored and are the focus of this current project.

The striatum is an inhibitory nucleus in the basal ganglia composed primarily of inhibitory projection neurons called medium spiny neurons (MSNs) that constitute the "direct pathway," and the D2-MSNs form the "indirect pathway." The balance between these two supposedly antagonistic pathways is considered critical in controlling movement (Kreitzer and Malenka, *Neuron 60:* 543 [2008]). We hypothesized that on learning the tonecloud task, changes in the strength of the cortical synapses onto neurons depend on which pathway it belongs to and whether it is tuned to low- or high-frequency sounds.

To test our hypothesis, we are using transgenic mouse lines that will allow us to measure the learning-induced changes in synaptic strength in the direct versus indirect pathway neurons. We have already established that the corticostriatal projections in the mice are arranged in a tonotopic manner as seen in rats. Moreover, training mice on the tonecloud task induces a specific pattern of plasticity that accurately reflects the tone-response association. Also, reversing the tone-response association faithfully reverses the plasticity pattern observed. We are now using the whole-cell patch-clamp method on acute slices prepared from brains of trained transgenic animals to measure the synaptic strength of the cortical afferents onto single neurons of each pathway. We believe these experiments will help us directly relate plasticity in striatal circuitry to a relevant behavioral output and therefore further our understanding of the role of striatal plasticity in acquiring decision-making behaviors.

Developing Synaptic PLA (Proximity Ligation Assay) as a Method to Detect Recently Potentiated Synapses

H. Zhan, S. Ghosh

The formation of memory and process of learning have long been attributed to changes in synaptic strength between specific cell populations in the brain. Yet, it remains highly challenging to identify which specific synapses undergo such changes when an animal acquires a new memory or learns to perform a certain task. Popular techniques such as staining for "immediate early genes" (e.g., *c-fos* and *arc*) can only indicate which cells in the brain are activated during a behavior but fail to provide answers at a subcellular resolution (i.e., at the level of a synapse). Two-photon imaging techniques may provide the necessary resolution, but they remain extremely low-throughput and technically challenging.

We are developing SYNPLA (SYNaptic proximity ligation assay), a sensitive, specific, and high-throughput method for detecting the synaptic plasticity between candidate neuronal populations. The most common form of synaptic plasticity is long-term potentiation (LTP), and decades of research have established that during LTP, GluA1 receptors move into synapses. Hence, SYNPLA exploits the PLA to detect synaptic colocalization of the pre- and postsynaptic proteins, neurexin and GluA1, respectively. SYNPLA results in greatly amplified, punctate fluorescent signal at recently potentiated synapses, allowing their easy and high-throughput identification using light microscopy.

We have shown SYNPLA in cultures of dissociated hippocampal neurons, and are now working to increase signal-to-noise and throughput. We are also working to optimize the SYNPLA protocol for detecting potentiated synapses in a rat brain in response to learning. Once perfected, this technique would allow a high-throughput identification of potentiated synapses in specific pathways or cell types of a rodent brain in response to complicated behaviors.

Using Transcriptional Information for Brain Sample Registration to a Reference Brain Atlas

S. Lu

Current methods of mapping brain areas from an experimental mouse to a second mouse or atlas depend primarily on anatomical landmarks. These landmarks, however, are often variable across animals and may be further distorted in animal models of neurodevelopmental and neuropsychiatric disorders. We propose to complement this anatomy-based method by integrating a second source of information based on transcriptomics. By looking at bulk transcriptomic data collected with spatial information from WT brain slices, we hope to find differences in transcriptional profiles across brain areas, which can then be used to aid registration of microdissected brain samples from MAPseq.

Mapping Neuronal Projections Using In Situ Barcode Sequencing with BARseq

X. Chen, Y.C. Sun, H. Zhan

We have been developing BARseq, a method of in situ sequencing optimized for RNA barcodes. Using this method, we were able to examine the spatial organization of neurons during multiplexed projection mapping. We confirmed known laminar organization of projection neuron classes in the mouse auditory cortex. We also identified subclasses of projection neurons that were mostly intermingled in their laminar positions.

Identification of Gene Expression Correlates of Long-Range Neuronal Connectivity

X. Chen, Y.C. Sun

The selective expression of genes controls many cellular properties, including neuronal connectivity, electrophysiology, and response properties to extracellular signaling. These different properties allow neurons to play diverse roles in neuronal circuits. Correlating gene expression with neuronal properties at cellular resolution is thus important in understanding the logic behind neuronal circuits. Although MAPseq allows multiplexed projection mapping at cellular resolution, such information cannot be combined with gene expression because of the loss of cellular identities. To correlate gene expression with projections, we developed a targeted in situ sequencing–based technique that reads out both barcodes and endogenous mRNAs. Using this technique, we are currently correlating neuronal projections from the mouse visual cortex to the expression of genes differentially expressed in different neuronal cell types.

High-Resolution Barcode Sequencing in Expanded Brain Slices

G. Henry

Recently, the Zador lab has made considerable progress in developing neuronal circuit–mapping technologies that rely on high-throughput sequencing of barcoded neurons. For example, the MAPseq method enables bulk sequencing of barcodes present at both a source and distant projection site, whereas BaristaSeq leverages Illumina sequencing chemistry to obtain barcode sequence in an in situ and nondestructive manner. In the past year, we have adapted the in situ sequencing scheme and combined it with expansion microscopy (ExM). Most super-resolution imaging methods rely on sophisticated microscopes that use advanced optics. Unlike these methods, ExM generates high-resolution images by isotropically enlarging or expanding tissue (3× –5×), which is then imaged by standard light microscopy. The increase in resolution is of interest to our efforts because it will enable simultaneous barcode sequencing and synapse identification.

Before brain slices are expanded, they are embedded in an acrylamide hydrogel. In our scheme, barcode-derived cDNA is anchored into the hydrogel network. In addition, oligo-coupled antibodies are used to memorialize the position of both synaptic proteins and GFP produced by the barcoded virus library, which uniquely marks each infected neuron. After the gel-anchoring steps, the sample is cleared and expanded. In the past year, we have (1) optimized barcode amplification from gel-anchored cDNA, (2) developed a novel hydrogel workflow that permits flexible gel expansion, and (3) worked through the details of merging barcode amplification with immunohistochemical detection of protein antigens (e.g., synaptic markers).

PUBLICATIONS

Marbach F, Zador AM. 2017. A self-initiated, two-alternative forced choice paradigm for head-fixed mice. bioRxiv doi: 10.1101/073783.

Peikon I, Kebschull J, Vagin V, Ravens D, Sun YC, Brouzes E, Corrêa I Jr, Bressan D, Zador A. 2017. Using high-throughput barcode sequencing to efficiently map connectomes. *Nucleic Acids Res* **45:** e115.

In Press

Chen X, Sun YC, Church GM, Lee JH, Zador AM. 2018. Efficient in situ barcode sequencing using padlock probe–based BaristaSeq. *Nucleic Acids Res* **46:** e22.

Han Y, Kebschull JM, Campbell RAA, Cowan D, Imhof F, Zador AM, Mrsic-Flogel TD. 2018. The logic of single-cell projections from visual cortex. *Nature* **556:** 51–55.

PLANT BIOLOGY

David Jackson and colleagues study genes and signals that regulate plant growth and architecture. They are investigating a unique way in which plant cells communicate, by transporting regulatory proteins via small channels called plasmodesmata. These channels, which direct the flow of nutrients and signals through growing tissues, are regulated during development. The team discovered a gene encoding a chaperonin, *CCT8*, that controls the transport of a transcription factor SHOOTMERISTEMLESS (STM) between cells in the plant stem cell niche, or meristem. STM is critical for stem cell maintenance, and studies of the *CCT8* gene indicate that movement of STM between cells is required for this function. The lab also continues to identify other genes that control plant architecture through effects on stem-cell maintenance and identity, and their work has implications for crop yields. Recent examples include discovery of a subunit of a heterotrimeric G protein that is conserved throughout animals and plants, and their studies indicate that this gene controls stem cell proliferation. They have found that in plants, the G protein interacts with a completely different class of receptors than in animals. Their discovery helps to explain how signaling from diverse receptors is achieved in plants. This year, they also showed that weak mutations in one of the receptor proteins can enhance seed production in maize, which could lead to yield increases. Separately, the lab has characterized system-wide networks of gene expression, using "next-gen" profiling and chromatin immunoprecipitation methods that have revealed many new hypotheses in developmental networks controlling inflorescence development. They are also developing a collection of maize lines that can drive expression of any reporter or experimental gene in any tissue type—tools of great interest to maize researchers that are being made available to the broader scientific community, enabling experiments never before possible in crop plants.

Zachary Lippman's research focuses on the process of flowering and flower production, which is a major contributor to plant reproductive success and agricultural yield. By identifying genes that control how tomato plants produce their flowers in their characteristic repeated zigzag arrangement (e.g., tomatoes on a vine), Lippman's lab is addressing when and how flowering branches, known as inflorescences, develop on plants, particularly fruit-bearing plants. Of particular interest is how these "reproductive phase transitions" have contributed to the evolution of diverse inflorescence branching patterns in tomato's larger Solanaceae family, which includes plants that make just one flower, such as pepper and petunia, in each inflorescence, to plants whose inflorescences produce dozens of branches and hundreds of flowers, such as many wild species of tomato. Using a combination of genetic, genomic, and molecular approaches, Lippman is dissecting the gene networks that are responsible for the variation in inflorescence branching found in nature. He hopes to leverage these discoveries to improve crop yields. Already, his work on genes responsible for the production and activity of a universal flowering hormone known as florigen has resulted in novel approaches to fine-tune plant architecture and flower production, boosting yield beyond leading commercial varieties. To continue hunting for new genes, Lippman has adopted a systems-biology approach and next-generation sequencing technology to capture those genes that are active as stem cells mature from a vegetative to a reproductive state. Nearly 4000 genes were found to reflect the existence of a "maturation clock," and one of the clock genes known as *Terminating Flower* acts as a key regulator to maintain a progressive pace to flowering—which, in turn, dictates how many flowers are produced on each tomato inflorescence. Finally, the Lippman lab determined the genome sequence of the "currant tomato," the wild ancestor of larger-fruited cultivated tomatoes, to better understand how flower and fruit production changed during the process of crop domestication.

Epigenetic mechanisms of gene regulation—chemical and conformational changes to DNA and the chromatin that bundles it—have had an important impact on genome organization and inheritance and on cell fate. These mechanisms are conserved in eukaryotes and provide an additional layer of information superimposed on the genetic code. **Robert Martienssen**, a pioneer in the study of epigenetics, investigates mechanisms involved in gene regulation and stem cell fate in yeast and model plants including *Arabidopsis* and maize. He and his colleagues have shed light on a phenomenon called position-effect variegation, caused by inactivation of a gene positioned near densely packed chromosomal material called heterochromatin. They have discovered that small RNA molecules arising from repeating genetic sequences program that heterochromatin. Martienssen and colleagues have described a remarkable process by which "companion cells" to sperm in plant pollen grains provide them with instructions that protect sperm DNA from transposon damage. They found that some of these instructions, or epigenetic marks, could be inherited in the next generation. These marks, and the small RNA that guide them, can sense the number of chromosomes inherited from pollen and may allow *Arabidopsis*, a flowering plant, to produce egg cells without meiosis—an important step toward a long-time goal of plant breeding: generating clonal offspring to perpetuate hybrid vigor. The lab has also shown that when RNA polymerase II has transcribed a stretch of DNA, the RNA interference mechanism causes the enzyme to release its hold on the DNA and fall away. This allows the replication fork to progress smoothly and the DNA strands to be copied; histone-modifying proteins, which follow right along, establish heterochromatin. Martienssen's group also continues to work on problems related to the creation of plant-based biofuels. As part of a collaborative project to generate a high-quality full-genome map of the oil palm plant, Martienssen and his colleagues identified a transposon whose modification controls the yield of oil palm trees. This discovery will increase yields and should lessen the environmental burden of oil palm production, which often threatens already endangered rainforest lands.

Plants and animals interact with their environment. Because plants are unable to move around, they are sensitive to their surrounding environment and modify their development according to external signals. Plants face variability in growth conditions—temperature, light quality and quantity, herbivores, pathogens, water availability, etc. Yet, plants respond to these biotic and abiotic factors and survive substantial fluctuations in their environment. Plants also must balance the range of potential threats and benefits confronting them and should make appropriate decisions on resource allocation. Such adaptability is essential for the sessile nature of the plants. The mechanisms that underlie this adaptability likely involve complex signaling to generate the appropriate response. In some adaptive responses, for example, when the plants have to cope with climate change and increased competition for light, there is a decrease in productivity (yield, biomass) as the plant reallocates resources to better adapt. The **Ullas Pedmale** lab seeks to determine the mechanisms behind how a plant perceives and successfully adapts to its environment. The team also aims to understand how a plant must integrate intrinsic and extrinsic cues and "decide" how best to respond to environmental cues. Understanding how plants deal with and respond to a multitude of environmental signals could help to develop crops that cope with unfavorable growth conditions without significant changes in yield.

The growing tips of plants, called meristems, contain a population of stem cells that serve as a persistent source of daughter cells from which new organs, such as leaves, arise. **Marja Timmermans** and colleagues are studying the genetic networks that regulate plant stem cell activity. Using genomic approaches, they have defined gene expression signatures that distinguish indeterminate stem cells from their differentiating derivatives. They have also worked out the mechanism that suppresses stem cell fate to allow cells to differentiate, and they have shown that this process requires a highly conserved epigenetic gene silencing mechanism. In particular, Timmermans' group has shown that specific DNA-binding proteins mediate the recruitment of Polycomb

repressive complexes to stem cell factors, an action that stably represses their expression in differentiating organs. This work addresses a major unresolved question in the field of epigenetics: How do Polycomb proteins, which do not bind DNA themselves, recognize defined targets? Plant stem cells also produce signals important for the patterning of lateral organs. The lab has discovered that small RNAs can traffic from cell to cell and are among the stem cell–derived signals. They have found that polarity in leaves is established via opposing gradients of mobile small RNAs that act as morphogen-like signals. Their most recent findings identified a third small RNA gradient involved in maintenance of organ polarity. These findings illustrate the complexity with which small RNAs generate developmental patterns. Currently, they are investigating parameters of small RNA mobility and the unique patterning properties of resulting small RNA gradients. Mathematical modeling predicts that such gradients might serve to generate robustness during development.

DEVELOPMENTAL BIOLOGY—STEM CELL SIGNALING AND CROP ARCHITECTURE

D. Jackson L. Brine, H. Claeys, P. Cunniff, E. Demesa-Arevalo, B.I. Je, E. Kallinos, M. Kitagawa, L. Liu, L. Marmorale, S. Pouzet, T. Skopelitis, Q. Wu, F. Xu, X. Xu, M. Yuan

Our research asks how the growth of plants is controlled, with the ultimate goal of improving crop yields. We identify the genes, signals, and pathways that regulate plant architecture and development. Like all organisms, plants develop by carefully controlling the movement of information molecules between cells during their growth. We are interested in discovering the signals that carry this information, in finding out how the signals are transmitted, and learning how they function.

A major focus has been identification of genes that control stem cell signaling pathways. In the past year we reported the discovery of a new player in stem cell control that acts downstream of the receptors we previously identified and helps explain how cells can decipher the many different signals they receive. We also described the biological relevance of cell to cell transport of a nuclear protein, SHOOTMERISTEMLESS, and are finding new roles for G proteins in stem cells. We continue to expand our use of CRISPR genome editing and to generate tools for the maize research community by creating a collection of lines that can drive expression in any tissue type. Such tools are of great interest to researchers and are being made available to the broader scientific community, enabling experiments never before possible in crop plants.

The Regulation of Meristem Size in Maize

Byoung Il Je, Qingyu Wu, Fang Xu [in collaboration with Robert Meeley, DuPont Crop Genetics]

All plant organs derive from populations of stem cells called meristems. These stem cells have two purposes: to divide and maintain themselves and to give rise to daughter cells, which will differentiate into plant organs. Consequently, meristems must precisely control the size of the stem cell niche via a network of positive and negative feedback signals. A loss of function in a negative regulator of stem cell fate can result in an enlarged or fasciated meristem phenotype and a dramatic alteration in the morphology of the maize ear and tassel. Maize is an excellent genetic model system because of a large collection of developmental mutants and a sequenced reference genome. Our laboratory has undertaken a forward genetic approach to identify key regulators of stem cell homeostasis and meristem size. Two previously cloned mutants, *fasciated ear2* and *thick-tassel dwarf1*, encode orthologs of the *Arabidopsis thaliana* genes *CLAVATA1* and *CLAVATA2*, indicating the well known *CLAVATA-WUSCHEL* regulatory feedback loop is conserved from dicots to monocots. However, little else is known about the control of this important developmental process in maize. Here, we describe progress in identifying additional genes contributing to stem cell homeostasis.

A common class of proteins that signal directly downstream of cell surface receptors is the heterotrimeric G protein, consisting of Gα, Gβ, and Gγ subunits. Our previous work found that the maize *COMPACT PLANT2* (*CT2*) gene, which encodes the α subunit of a heterotrimeric GTPase (Gα), functions in the CLAVATA pathway to control meristem size through its interaction with FEA2. To understand the mechanism of CT2 signaling in meristem development, we generated a constitutively active (GTPase-dead) version of CT2 ($CT2^{CA}$) by introducing the Q223L mutation in the context of a native construct. To ask if this mutation abolished GTPase activity, we performed in vitro GTP-binding and GTPase activity assays using fluorescent BODIPY-GTP, where an increase in fluorescence over time corresponds to GTP binding, and a subsequent decrease corresponds to GTP hydrolysis. We found that the $CT2^{CA}$ protein had similar GTP binding as CT2, but lacked GTPase

activity. We further tested whether $CT2^{CA}$ interacted with Gβγ in a yeast-three-hybrid (Y3H) system. In contrast to CT2, we found that $CT2^{CA}$ did not interact with the Gβγ dimer. Both results confirmed that $CT2^{CA}$ is indeed constitutively active and that it could no longer form a heterotrimeric complex with Gβγ. Expression of *CT2^CA^* in a *ct2* mutant background resulted in interesting phenotypes, including higher spikelet density and kernel row number, larger shoot apical and ear inflorescence meristems, and smaller leaf angles compared with normal sibs. These results suggest that manipulation of the GTPase activity of maize Gα subunit is a promising approach to improving agronomic traits of maize.

In addition to CT2, the sole canonical Gα of maize, there are three "extra large G proteins" (XLGs), which have a domain with homology to Gα as well as additional domains. We first demonstrated that the three ZmXLGs function in a heterotrimeric G protein complex by showing their interaction with a Gβγ dimer in the Y3H system. Then we used CRISPR-Cas9 to create different alleles of all three *ZmXLG*s. Knocking out each single *ZmXLG* did not alter development, whereas knocking out any two *ZmXLG*s led to a modest but significant reduction in plant height, but did not affect SAM size, indicating that loss of any two *ZmXLG*s can be partially compensated by other *XLG*s or by *CT2*. We also showed that mutation of any two *ZmXLG*s in the *ct2* mutants dramatically enhanced their dwarf and enlarged SAM phenotypes. In contrast, although both *CT2* and *ZmXLG*s are expressed in the maize inflorescence, *ZmXLG* knockouts did not enhance the *ct2* ear fasciation phenotype, suggesting that CT2 is the major Gα functioning in inflorescence meristem development. Surprisingly, all triple *Zmxlg* mutant plants showed a striking arrest phenotype, as the plants died at the seedling stage.

We also knocked out the sole *Gβ* gene (*ZmGB1*) of maize, again using CRISPR-Cas9. We found that the *Zmgb1* null plants also died at a very early stage of seedling development. In addition, they showed overaccumulation of H_2O_2 and salicylic acid, constitutive activation of MAP kinases, and up-regulation of *PR1 (PATHOGENESIS-RELATED1)* and *PR5*, two immune marker genes. These results suggest that *ZmGB1* mutation causes autoimmune symptoms. Because the mutants died at a very early stage, we were unable to study shoot meristem and inflorescence phenotypes. Therefore, we introgressed *Zmgb1* null mutants into different lines to see if the lethal phenotype can be suppressed, and we indeed found that it was partially suppressed in certain genetic backgrounds and could grow to the reproductive stage (Fig. 1A). In the lethality-suppressed background, the *Zmgb1* mutants had significantly larger SAMs and fasciated inflorescence meristems in both ear and tassel (Fig. 1B,C). Concurrently, by map-based cloning of a fasciated mutant, *tr183,* we identified a second allele of *Zmgb1*, which contained an aspartic acid to asparagine amino acid change in a conserved residue of the WD40 domain. *Zmgb1-tr183* mutants have fasciated ears and thick tassels and are dwarf with enlarged SAMs. They also develop lesions, with sporadic cell death and immune genes induced, like the *Zmgb1* null mutants. *Zmgb1-tr183* failed to complement the *Zmgb1-cr mutants*, suggesting that they are allelic. Further studies will focus on understanding the roles of Gβ and XLGs in maize development and immune responses.

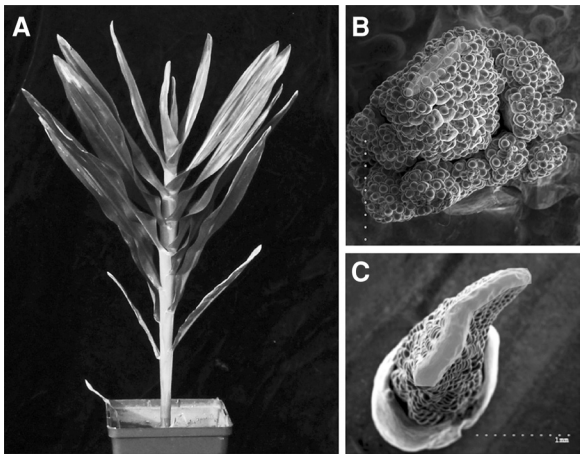


Figure 1. The lethality of *Zmgb1-cr* mutants is suppressed in certain genetic backgrounds. (*A*) The *Zmgb1-cr* mutants in the CML103 background can grow to the reproductive stage, and tassel (*B*) and ear (*C*) inflorescence meristems are fasciated.

Another fasciated ear mutant that we cloned is *fasciated ear 3* (*fea3*). *FEA3* encodes a predicted leucine rich repeat receptor-like protein, related to *FEA2*. In situ hybridization and imaging of RFP-tagged transgenic plants show that FEA3 is expressed in the organizing center of the SAM and in leaf primordia. Remarkably, expression of maize WUSCHEL, a marker for the stem cell niche organizing cells, spreads downward in *fea3* mutants, which is strikingly different from its response in the known *CLAVATA* stem cell mutants.

FEA3 responds to a newly discovered CLV3-related (CLE) peptide, ZmFCP1 (FLORAL ORGAN

NUMBER2-like CLE protein 1), suggesting that it functions in a new pathway for stem cell control that is spatially distinct from the known CLV receptors. We also found that *FEA3* and *FCP1* homologs function in meristem control in *Arabidopsis*, suggesting this new pathway is universal in flowering plants. Meanwhile, to further understand the role of FEA3 in stem cell signaling, IP-MS (immunoprecipitation-mass spectrometry) with FEA3-tagged transgenic plants will be used to find interacting proteins. Plants carrying FEA3 fused with amino-terminal or carboxy-terminal tandem affinity purification (TAP) tags (YFP-STREP tag) have been constructed and will be used for tandem affinity purification followed by MS to isolate FEA3 interactors. As a backup, FEA3-tagged lines were also crossed to a *branched silkless; Tunicate* (*bd-Tu*) double mutant, which makes a mass of proliferating meristems and results in strong enrichment of developmental protein complexes. In addition, CRISPR technology is being used to knock out FEA3 homologs to study their function in maize. We already obtained different CRISPR alleles for four FEA3 homologs; however, no obvious phenotype was observed, and by more careful phylogenetic analysis it appears that additional homologs may need to be mutated to overcome this redundancy.

Weak alleles of fasciated ear mutants can improve maize yield traits, such as kernel row number, by increasing meristem size and number of primordia while maintaining structural integrity of the meristem. We found that *fea3* weak allele hybrids also enhance yield traits as well as overall yield. These results are particularly exciting, because in our previous studies of weak *fea2* alleles we found an increase in kernel row number but no overall increase in ear weight (due to a compensatory reduction in kernel size). Therefore, the newly identified *FEA3* signaling pathway could be used to develop new alleles for crop improvement.

Having found multiple receptors and CLE peptide ligands, we were intrigued by how specificity is obtained. Using peptide assays, we found that FEA2 also participates in ZmFCP1 signaling, but CT2 does not. CORYNE (CRN, a membrane, localized pseudokinase) is a CLV2 interactor first identified in *Arabidopsis*. We identified two *Zmcrn* mutants by reverse and forward genetic approaches and found that they both have a fasciated ear phenotype and enlarged meristems. Double mutant analysis showed that *fea2* is epistatic to *Zmcrn*, and ZmCRN interacts physically with FEA2, indicating that they function together and ZmCRN might be a downstream signaling component. We previously found that CT2, the G protein α subunit, also functions with FEA2, revealing CT2 as another signaling component. *Zmcrn ct2* double mutants had additively enlarged SAM size and enhanced fasciated ears compared to single mutants, suggesting that ZmCRN and CT2 do not function in the same pathway. Using peptide response assays, we also discovered that FEA2 transmits signals from two distinct CLE peptides, the maize CLV3 ortholog ZmCLE7 and ZmFCP1, through the two different downstream signaling components, CT2 and ZmCRN. Our data provide a novel framework to understand how diverse signaling peptides can activate different downstream pathways through common receptor proteins and bring new insights to understand signaling specificity by different ligands. In addition, we found that *ZmCRN* is significantly associated with kernel row number (KRN) in a maize association panel, and mutations in ZmCRN lead to enlarged ear meristems, suggesting it is a good candidate for manipulation to improve maize yield.

Genetic Redundancy in Circuits Controlling Meristem Development

L. Liu, E. Demesa-Arevalo, B.I. Je, F. Xu, T. Skopelitis
[in collaboration with M. Bartlett, University of Massachusetts Amherst; Z. Nimchuck, University of North Carolina at Chapel Hill; B. Yang, Iowa State University; Z. Lippman, CSHL]

The CLE (CLAVATA3/endosperm surrounding region–related) peptides are fundamental players in meristem maintenance in plants, as they are mobile signals that establish feedback circuit signaling between stem cell differentiation and division. Disruption of this pathway generates overproliferation in meristems, or fasciation. In our laboratory, we have described different fasciated mutants encoding leucine-rich repeat receptor-like kinases (LRR-RLKs) or receptor-like proteins (RLPs); however, the signals perceived by many of these receptors remain elusive. Genetic evidence in maize suggests a divergence in signaling pathways controlling meristem size. The putative orthologs in maize for CLAVATA1 (THICK TASSEL DWARF1, TD1), CLAVATA2 (FASCIATED EAR2, FEA2), and FASCIATED EAR3 (FEA3) have synergistic effects, suggesting additional pathways control

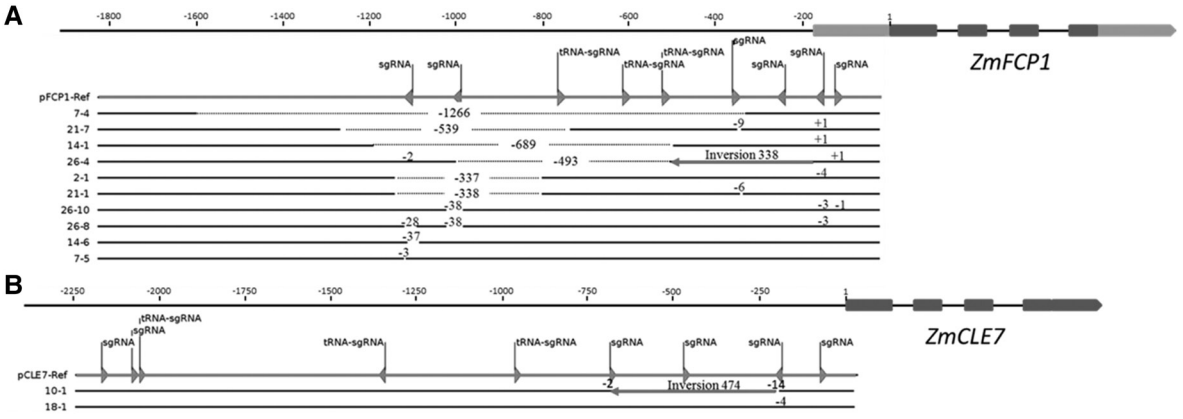


Figure 2. CRISPR mutagenesis of promoters of *ZmFCP1* (*A*) and *ZmCLE7* (*B*).

meristem size. Recently 49 CLE peptide genes were identified in maize, suggesting either specialization or redundancy in these ligands.

To analyze the role of CLE peptide ligands in maize meristem regulation and their involvement in redundant circuits, we are generating CRISPR knockouts. We first analyzed expression patterns from publicly available data sets combined with our transcriptional profiles from meristematic tissues and identified 31 candidates expressed in SAM and inflorescence meristems. Up to now, we have made CRISPR alleles for approximately 24 of them. Mutant alleles for ZmCLE7 and ZmFCP1 show that they have a major role in meristem regulation. To identify additional *Zm*CLE peptides that regulate meristem size in a redundant way, we are generating CRISPR mutants in other ZmCLE genes and crossing them to make higher-order mutants, especially for those expressed in the same tissues or phylogenetically related. ZmCLE14 is most closely related to ZmCLE7 and ZmFCP1; we recovered four mutant alleles in this gene, and we will analyze its genetic redundancy and developmental roles in meristem regulation.

In *Arabidopsis*, *clv1* mutant phenotypes are enhanced by mutations in the related RLKs *BARELY ANY MERISTEM 1*, *2*, and *3* (*BAM1*, *BAM2*, and *BAM3*). In maize, we identified seven *BAM*-like genes and a single *CLAVATA1* gene (*TD1*). To characterize their function, we used CRISPR-Cas9 to mutate all of them. The genetic interactions between different BAM genes and CLEs will allow us to dissect additional signaling pathways for meristem maintenance in maize.

Weak alleles of *fea2* and *fea3* have great potential for yield improvement by increasing kernel row number. Therefore, we are also using CRISPR to create mutations in CLE gene promoters, focusing on those that influence ear development, to create weak alleles for enhancement of yield. The promoters (~2 kb) of *ZmFCP1* and *ZmCLE7* were targeted by multiplex gRNA (eight to nine sgRNAs) CRISPR, and a variety of edited haplotypes was obtained (Fig. 2). Phenotype analysis will evaluate the effect of these mutations.

Control of Shoot Branching and Determinacy

H. Claeys, E. Demesa-Arevalo, X. Xu, S. Pouzet, M. Yuan, T. Skopelitis [in collaboration with B. Yang, Iowa State University]

The *RAMOSA* (*RA*) genes in maize function to impose determinacy on axillary meristem growth; consequently, *ra* loss-of-function mutants (*ra1*, *ra2*, and *ra3*) have more highly branched inflorescences. *RA3* encodes a trehalose phosphate phosphatase, which catalyzes the conversion of trehalose-6-phosphate (T6P) to trehalose. T6P is an important regulatory metabolite that connects sucrose levels, and thus the sugar status, to plant growth and development, but its mode of action is still unclear. *RA3* is expressed at the base of axillary inflorescence meristems and localizes to distinct puncta in both nuclear and cytoplasmic compartments, suggesting that its effect on development may not be simply metabolic. These data support the hypothesis that *RA* genes may serve as mediators of signals, maybe a sugar signal, originated

at the boundary domain and regulating determinacy. *RA3* itself may have a transcriptional regulatory function, because it affects the expression of specific genes.

We are taking genetic approaches to identify factors that act in the same pathway by screening for enhancers of the *ra3* phenotype. Typically, *ra3* mutants have three to eight branches only at the base of the ear. We mutagenized *ra3* and looked for plants that have more branches and/or have branches in the upper part of the ear. Twenty mutants have been isolated so far, and mapping is under way.

So far, four independent mutant alleles of *TPP4* (*TREHALOSE-6-PHOSPHATE PHOSPHATASE*), a *RA3* paralog, were identified (Fig. 3), and we confirmed that *TPP4* is the causative gene using additional CRISPR-Cas9-generated alleles. *TPP4* is present in the same expression domain as *RA3* and is normally only modestly expressed, but upon mutation of *ra3* its expression is up-regulated, suggesting that it acts as a partially redundant backup for *RA3* (see Fig. 4). All EMS-induced alleles contain single amino acid substitutions, and some of the resulting mutant proteins still have considerable residual enzymatic activity despite all alleles having similar phenotypic strength. This indicates that there is no straightforward relationship between TPP activity and phenotype, and that additional regulatory functions of TPP4 may be important. Using CRISPR-Cas9, we also generated knockouts of *TPP12*, a more distant family member that is also expressed in developing inflorescences, but unlike *TPP4*, mutating *TPP12* does not enhance the RA3 phenotype, showing functional divergence within the TPP family.

Another *ra3* enhancer mutant was mapped to a gene encoding an RNA-binding protein that is known to play a role in inflorescence development. We have made maize lines carrying functional YFP fusions and will use these to further characterize the role of this protein in meristem determinacy, and its relationship with RA3, by looking at protein–protein and protein–RNA interactions. For the other *ra3* enhancer mutants, we are currently at various stages of mapping and confirming the causal genes.

In a parallel approach to understand how RA3 functions, we are screening for interacting proteins. To this end, a yeast two-hybrid screen was performed, and a number of predominantly nuclear proteins were found to interact with RA3, which fits with its partial nuclear localization. In parallel, ear primordia from plants expressing a Flag-HA-tagged RA3 protein were used for IP-MS, revealing a number of potential in vivo interactors. The biological roles of a number of the interactors are currently being studied using insertional mutants, and mutant alleles are being generated using CRISPR-Cas9. We analyzed CRISPR alleles in two families of RA3 interactors—the RNA recognition motif (RRM) *Zea mays* SCAFFOLD ATTACHMENT FACTOR B (*Zm*SAFB) and ZmSAFB-like and the vascular plant one-zinc-finger (*Zm*VOZ) transcription factors—to analyze their genetic interactions with *ra3*.

So far, two consecutive field trials showed *Zmsafb ra3* mutants have around two times more branches in ear and tassels compared with the single *ra3* mutants, suggesting *SAFB* functions with *RA3* to regulate determinacy. Additionally, we generated mRFP1-*Zm*SAFB,

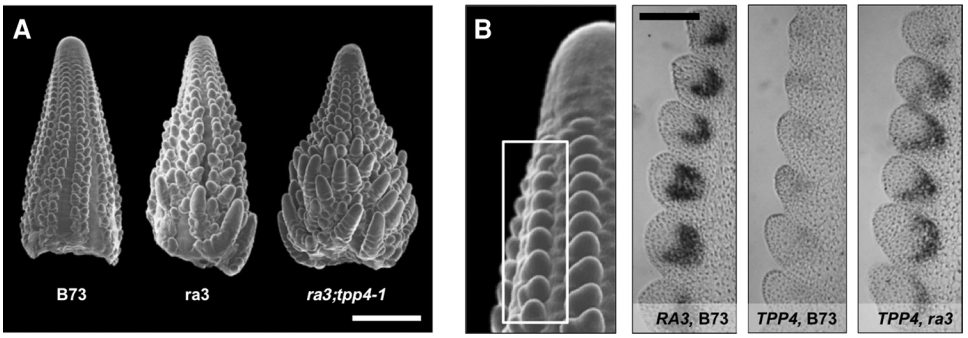


Figure 3. Mutations in *TPP4* enhance the *ra3* phenotype. (*A*) SEM pictures of ears from B73, *ra3*, and *ra3;tpp4-1* plants, showing the presence of a few basal branches in *ra3* ears, which are absent from B73 and greatly increased in *ra3;tpp4-1*. Scale bar, 1 mm. (*B*) Expression of RA3 and TPP4 at the base of spikelet pair meristems, as shown by in situ hybridization. The region shown is indicated on a SEM picture in the *left* panel. Scale bar, 100 µm.

ZmVOZ1-RFP, and YFP-ZmVOZ5 transgenic lines to characterize localization patterns in ear development of these candidate interactors. We are also using these lines to test their in vivo interactions with RA3. We optimized protocols to immunoprecipitate (IP) mRFP1-ZmSAFB, and we are introducing our three FP-tagged lines in the bd;Tu; FLAG-HA-tagged RA3 to perform co-IP and also to identify novel protein–protein or protein–DNA/RNA interactions involved.

We are also combining a number of approaches to directly understand the contribution of the enzymatic function of RA3 to its biological function. We have achieved partial complementation of the *ra3* mutant with a catalytically dead mutant version of RA3 and are now refining this experiment by mutating the endogenous *RA3* to a catalytically dead version using CRISPR-Cas9-mediated base editing. We are also using CRISPR-Cas9 to generate mutants in TREHALOSE PHOSPHATE SYNTHASES (TPSs), which catalyze the step preceding RA3 in the trehalose metabolic pathway. If accumulation of T6P is causing the *ra3* phenotype, then combining *ra3* with *tps* mutants should alleviate the phenotype, giving important insights into its mechanism.

Last, we are attempting to CRISPR the entire TPP gene family in *Arabidopsis*. So far, mutation of two RA3 homologs, *TPPI* and *TPPJ*, did not reveal any obvious phenotype, and we are now mutating additional family members to overcome the likely redundancy in this gene family.

Natural Variation in Inflorescence Architecture

H. Claeys, Q. Wu

Maize inflorescence architecture has been a target for extensive selection by breeders since domestication; hence, different maize inbreds vary greatly in these traits. The genetic basis underlying this diversity is largely unknown, but is of great interest for both fundamental and applied science. To identify natural variation relevant to inflorescence traits, we looked for inbred backgrounds that can enhance or suppress the phenotypes of different mutants. We focused on the 25 NAM (nested association mapping) founder inbreds because they were selected to capture the diversity of maize germplasm, and because of the genetic tools available for these lines. We crossed these 25 inbreds to our collection of mutants (in a B73 background) and screened the F2 for mutants with suppressed or enhanced phenotypes compared to the B73 background.

fea2-0 was strongly enhanced in the NC350 background. We mapped this enhancement to a single major-effect locus on chromosome 5 using both bulked segregant analysis and crosses to NC350/B73 recombinant inbred lines (RILs). The region we identified contains *THICK TASSEL DWARF1* (*TD1*), a gene that is known to be involved in meristem size regulation, and a *td1* loss-of-function mutant cannot rescue the enhanced phenotype, suggesting that *TD1* is the causal gene. Interestingly, we have also been able to demonstrate, using B73-NC350 heterogeneous inbred families (HIFs), that the NC350 allele of *TD1* positively affects KRN, demonstrating its usefulness in plant breeding. This confirms our laboratory's recent findings that hypomorphic alleles of genes involved in meristem size regulation lead to increased KRN, linking fasciation and KRN.

We also screened for natural modifiers of *ct2* and found it was dramatically enhanced in NC350 and HP301 backgrounds and suppressed in CML69. In the enhanced plants, the ear inflorescence meristems were extremely fasciated and had multiple branches. The CML69 background suppressed only the fasciated ear phenotype, whereas other *ct2* phenotypes, such as dwarf, wide and erect leaves, and dense spikelets, were not affected, indicating the suppressor may function specifically in ear inflorescence meristem regulation. Bulk segregant mapping analysis suggests that multiple loci contribute, and we are backcrossing modified plants to dissect the *ct2* modifier loci.

The Effects of Drought on Early Inflorescence Development

H. Claeys [in collaboration with H. Cline, B. Meeley, DuPont Crop Genetics; E. Vollbrecht, Iowa State University; S. Hake, USDA-UC Berkeley; J. Dinneny, Stanford University; A. Eveland, Danforth Center]

Drought stress is one of the major environmental factors limiting maize yield. Some progress has been made in studying how drought affects grain filling during the later stages of reproductive development, but almost nothing is known about how drought affects early inflorescence development, during which the number of grains is determined. It has been found that yield

is significantly affected when maize is subjected to drought at this stage. Therefore, we set out to address this question in collaboration with DuPont-Pioneer.

Last year, we completed a large pilot experiment in which we profiled ear tips from plants grown in automated greenhouses at a Pioneer facility in Iowa. Water was withheld from half of the plants during early ear development. This revealed a large transcriptional reprogramming of inflorescence meristems in response to drought, affecting many known developmental regulators. We identified the gibberellin and trehalose-6-phosphate (T6P) pathways as potential mediators of this response, and preliminary follow-up experiments have shown that ears of *ra3* mutants, which are affected in T6P metabolism, are more affected by drought. We are currently performing a second large transcriptome experiment to identify early signaling events after drought with greater developmental resolution. We have also identified inbreds with different sensitivity of ear growth to drought, which we plan to profile in the second phase of our drought experiments.

Generation of a Transactivation System for Functional Studies in Maize

E. Demesa-Arevalo, Q. Wu, B.I. Je, T. Skopelitis
[in collaboration with A. Chan, J. Craig Venter Institute; A. Sylvester, University of Wyoming]

Despite the abundance of resources available for the study of various model organisms, the classical approach of characterizing single-gene mutants still provides invaluable information regarding gene function. Using our experience in promoter analysis, we developed a pOp-LhG4 transactivation system in maize that allows us to express proteins in specific tissues and/or transient stages during development. The pOp-LhG4 system includes (1) a transcription activator, LhG4, which is a fusion between a high-affinity DNA-binding mutant of lac repressor, *LacI*^{His17}, and transcription-activation-domain-II of GAL4 from *Saccharomyces cerevisiae*, and (2) a chimeric promoter, pOp, that consists of *lac* operators cloned upstream of a minimal CaMV promoter not activated in the reporter lines until crossed with the LhG4 activator line. We have already generated 35 LhG4 drivers, as well as eight pOp responder lines. We confirmed the reliability of our system by crossing pOp::ZCN8 plants with a constitutive promoter line, pEF1A::LhG4, phenocopying the early flowering phenotype of ZCN8 overexpression. This system has been used to test new hypotheses in meristem regulation by differentiated cells; expressing pOp::*Zm*FCP1 using a leaf primordia-specific driver line, pYABBY14::LhG4, strongly inhibited meristem growth, confirming this feedback regulation. We are taking advantage of this system to analyze the overexpression of other meristem regulators. We generated a pOp::ZmCLE7 responder line and in contrast to overexpression of ZmFCP1, which suppressed development, ZmCLE7 did not show an effect in inflorescence meristems. However, when we used a pWUSCHEL1::LhG4 line to drive pOp::ZmCLE7 overexpression, we found axillary meristem suppression. These findings suggest specialized signaling networks controlling meristem architecture in maize (Fig. 4).

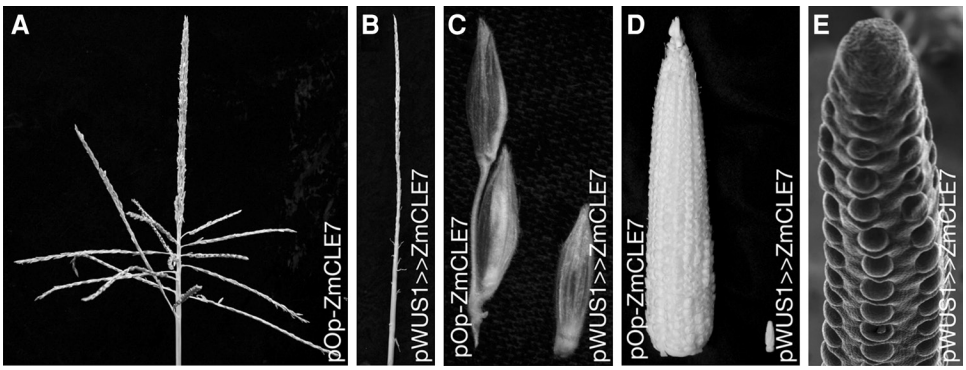


Figure 4. Overexpression of ZmCLE7 under pWUSCHEL1, using a two-component transactivation system, reduces tassel branching (*A*,*B*) and suppresses paired spikelets (*C*,*E*; ear SEM picture); and the ear inflorescence meristem is prematurely terminated, generating small ears in pWUS1 > >ZmCLE7 plants (*D*).

Currently we have close to 200 transgenic lines, including translational fusions, promoter fusions, promoter activator lines (LhG4), and responder lines (pOp). Data on the characterization of our FP and LhG4 transgenic lines, including confocal micrographs, movies, and recent publications, can be found on our website http://maize.jcvi.org/cellgenomics; seed are also available through the Maize Genetic Cooperation Stock Center.

Mechanism of Active Transport of Transcription Factors through Plasmodesmata

M. Kitagawa, P. Cunniff, T. Skopelitis

In plants, certain transcription factors (TFs) are actively and selectively transported between cells to target cells to specify their fates. These TFs are transported through plasmodesmata (PD), membrane-lined channels traversing the cell wall. To this date, however, the mechanism underlying the active and selective transport of TFs through PD has been largely unknown. Previously, we established a system for evaluating the capacity of the active transport of TFs in *Arabidopsis* seedlings using a mobile homeodomain TF, KNOTTED1 (KN1). Using this system, we isolated mutants in which the active transport of TFs may be defective. One such mutant encodes an aspartyl-transfer RNA synthetase (asp-RS), an enzyme that attaches L-aspartate onto its transfer RNA (tRNA) to function in translation. Interestingly, we found enlarged shoot apical meristem (SAM) and fasciation of stems in this mutant, suggesting this asp-RS regulates SAM development. Next, we observed the expression domain of *WUSCHEL* (*WUS*), which is a marker gene of the organizing center (OC) of the SAM, and found that the OC was expanded or split in the enlarged mutant SAMs, suggesting that this asp-RS regulates SAM organization. We next examined the subcellular localization of this asp-RS by fluorescent protein fusions and found a punctate pattern, implying that this asp-RS is located in specific organelles.

We also isolated two new mutants, *rb31-7* and *mk5-140*, by the screen described above. In the past year, we found using next-generation sequencing (NGS) mapping that the causal gene of both mutants is *At2g17510* (Fig. 5). This gene encodes ribosomal RNA processing protein (RRP) 44A, a subunit of the

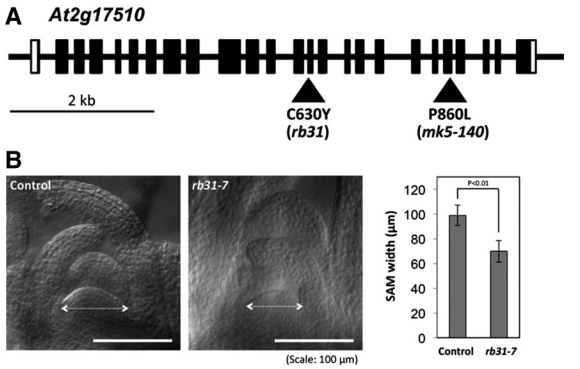


Figure 5. Characterization of PD trafficking mutants *rb31-7* and *mk5-140*. (*A*) Gene model of *At2g17510*, which is the causal gene of both of *rb31-7* and *mk5-140*. (*B*) SAM size is significantly decreased in *rb31-7*.

RNA exosome complex that functions in degradation and processing of RNAs, suggesting RNA degradation or processing by AtRRP44A may be involved in PD trafficking of mobile TFs. We also found that *rb31-7* has smaller SAMs (Fig. 5) and fusions between leaf and stem. These phenotypes are also observed in plants in which the active transport of the *Arabidopsis* KN1 homolog SHOOTMERISTEMLESS (STM) is artificially restricted (nonmobile STM plant), implying that the active transport of particular TFs including STM may be defective in *rb31-7*. Additionally, the decreased SAM activity in the *stm* weak allele, *stm-10*, could be enhanced in the *rb31-7 stm-10* double mutant. Thus, our results suggest that AtRRP44A is involved in SAM activity and organ separation by interacting with STM function and may regulate STM trafficking via PD, which is crucial for SAM regulation.

Functional Annotation of the Maize Genome by ChIP-seq and FACS

X. Xu, X. Wang, L. Wang, J. Drenkow [in collaboration with C. Ortiz Ramirez, K. Birnbaum, Center for Genomics and Systems Biology, New York University; T. Gingeras, D. Ware, D. Jackson, CSHL]

To develop functional annotation of the maize genome, we are conducting genome-wide transcription factor (TF) binding analysis by chromatin immunoprecipitation sequencing (ChIP-seq) and profiling B73 cell types by fluorescence-activated cell sorting (FACS). These are important goals of the MaizeCODE project, an initial analysis of functional elements in the maize genome.

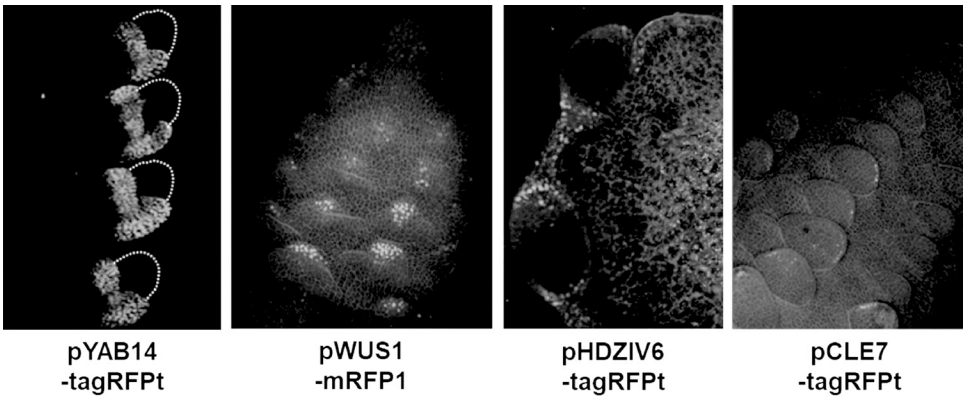


Figure 6. Fluorescent protein (FP)-tagged promoter lines marking different cell types or domains in maize ears using laser scanning microscopy.

To perform genome-wide TF binding site analysis, we have generated fluorescent protein (FP)-tagged transgenic lines for diverse TFs that function in different aspects of maize development. These TFs belong to several different families, such as the MADS box, TUNICATE1A, functioning in maize domestication, the homeodomain TF, WUSCHEL1 (WUS1), functioning in meristem maintenance, and the GATA TF, TASSELSHEATH1 (TSH1), functioning in bract suppression. To overcome limited tissue availability for conducting ChIP-seq, we crossed these lines into a double mutant, *branched silkless*;*Tunicate* (*bd*;*Tu*), that transforms the maize ear into a "cauliflower" with overproliferating meristems.

To profile cell types by FACS, we generated FP promoter lines to isolate specific cell types or domains for RNA-Seq, small RNA, DNA methylation, and ATAC-seq. The promoter lines cover several different types of cells or tissues in the meristem, such as organizing center (p*ZmWUSCHEL1*-mRFP1), lateral organ zone (p*ZmYABBY14*-tagRFPt), and epidermal zone (p*ZmHOMEODOMAIN-LEUCINE ZIPPER IV6*-tagRFPt) (Fig. 6). We also crossed these lines into *bd;Tu* double mutants to generate tissue for FACS. The specific expression domains of these reporter lines were confirmed by microscopy before sorting. Data from ChIP-seq and FACS profiling experiments are being generated and will be available through the CyVerse MaizeCODE project portal (in development) as a public resource.

PUBLICATIONS

Balkunde R, Kitagawa M, Xu XM, Wang J, Jackson D. 2017. SHOOT MERISTEMLESS trafficking controls axillary meristem formation, meristem size and organ boundaries in *Arabidopsis*. *Plant J* **90:** 435–446.

Jackson D. 2017. No sex please, we're (in)breeding. *EMBO J* **36:** 703–704.

Yang N, Xu X-W, Wang R-R, Peng W-L, Cai L, Song J-M, Li W, Luo X, Niu L, Wang Y, et al. 2017. Contributions of *Zea mays* subspecies *mexicana* haplotypes to modern maize. *Nat Commun* **8:** 1874.

In Press

Je BI, Xu F, Wu Q, Liu L, Meeley R, Gallagher JP, Corcilius L, Payne RJ, Bartlett ME, Jackson D. 2018. The CLAVATA receptor FASCIATED EAR2 responds to distinct CLE peptides by signaling through two downstream effectors. *eLife* **15:** e35673

Wu Q, Jackson D. 2018. Detection of MAPK3/6 phosphorylation during hypersensitive response (HR)-associated programmed cell death in plants. *Meth Mol Biol* **1743:** 153–161.

Xu F, Jackson D. Learning from CIK plants. *Nat Plants* **4:** 195–196.

MECHANISMS OF EPISTASIS AND QUANTITATIVE VARIATION IN PLANTS

Z.B. Lippman J. Dalrymple A. Hendelman E. Joshua A. Krainer C-T. Kwon Z. Lemmon S. Qiao D. Rodriguez-Leal S. Soyk X. Wang C. Xu

Plant reproductive success in nature and crop yield in agriculture rely on flowers, which are the foundation for fruit and seed production. Flowers develop from inflorescences, reproductive branching systems that originate from leaf-producing vegetative shoots when environmental and endogenous signals induce small populations of stem cells at the tips of shoots known as meristems to transition into reproductive states. The number of inflorescences produced on a plant, as well as how many branches and flowers form on each inflorescence, can vary dramatically both within and between species. At the center of this diversity lie two critical processes of stem-cell regulation that are the focus of research in the lab: (i) maturation, during which stem cells transition from a vegetative to a reproductive growth program, and (ii) proliferation, which controls stem-cell population size. Our research program integrates development, genetics, genomics, and gene editing to explore mechanisms of meristem maturation and maintenance and their relationship to plant shape, flowering, and flower production. We take advantage of extensive natural and mutant variation in inflorescence production and architecture in tomato and related nightshades to explore how differences in these processes explain the remarkable diversity in inflorescence production and complexity observed in nature and agriculture. Recent discoveries on these topics have led us to broader questions on the significance of genomic structural variation, gene redundancy, and epistasis in development, domestication, and breeding. Based on our fundamental discoveries, we are developing and applying innovative concepts and tools for crop improvement.

Revealing and Disarming Negative Epistasis on Yield in Tomato

S. Soyk, Z. Lemmon

The agricultural product of most major crops is derived from inflorescences. Depending on the plant species, reproductive meristems terminate immediately in a flower or give rise to a variable number of new inflorescence meristems that become additional flowers or flower-bearing branches. In tomato, a new inflorescence meristem emerges at the flank of each previous meristem and several reiterations of this process lead to inflorescences with multiple flowers arranged in a zigzag pattern, resulting in the familiar "tomatoes on the vine." During domestication and improvement of many crops, humans frequently selected for branched inflorescences with more flowers and fruits. However, the inflorescence of cultivated tomato has changed little from that of wild ancestors.

In a recent study, we discovered that rare tomato varieties with branched inflorescences from a large diversity panel carry mutations in two related MADS-box transcription factor genes (Fig. 1). By analyzing allele distributions across the germplasm, we found that one mutation arose during domestication of the berry-sized wild relative of tomato and enlarged the leaf-like organs on fruits. The second mutation was discovered in the last century and eliminated the abscission zone on fruit stems, providing the "jointless" trait that reduced fruit drop and postharvest fruit damage. However, because of an epistatic relationship between the two MADS-box genes, combining both mutations during breeding resulted in excessive inflorescence branching and reduced fertility. By dissecting and reorganizing these MADS-box mutations and a gene-edited family member, we developed strategies for fine-tuning MADS-box gene dosage to achieve weakly branched inflorescences with higher flower production and fruit yields. Notably, although breeders have used the jointless trait since the 1950s, our work identified the causative MADS-box gene mutation, which we have shown can now be easily engineered in elite breeding germplasm using CRISPR technology. This example of negative epistasis on yield from two beneficial mutations may be a microcosm for what is likely a multitude of both positive and negative allelic collisions that shaped crop domestication, improvement, and breeding but still await discovery.

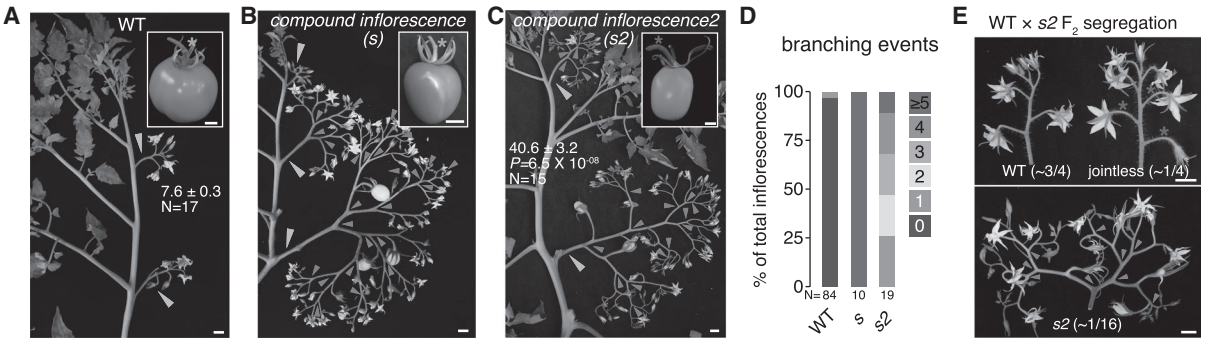


Figure 1. The *s2* inflorescence architecture variant branches because of delayed meristem maturation. (*A*) Typical wild-type (WT) tomato plant with unbranched, multiflowered inflorescences and jointed pedicels (asterisk in *inset*). Numbers in *A*–*C* indicate flowers per inflorescence (mean ± SEM, *N* = number of inflorescences). Arrowheads indicate successive inflorescences. *P*: two-tailed, two-sample *t*-test compared with WT. (*B*) The highly branched inflorescences and jointed pedicels of *s* mutants. Arrowheads indicate branch points. (*C*) *s2* mutant with moderately branched inflorescences and jointless pedicels (asterisk). (*D*) Quantification of inflorescence branching events in WT, *s*, and *s2*. (*E*) Phenotypic classes in a WT × *s2* F_2 population. The segregation ratio for the jointless pedicel phenotype and the branched inflorescence phenotype (*s2*) is given. Asterisks mark jointless pedicels. Scale bars, 1 cm. Further information available in Soyk et al. (2017).

Redundancy and Epistasis in the Control of Stem-Cell Proliferation and Meristem Size

D. Rodriguez-Leal, C-T. Kwon, A. Hendelman, C. Xu

The *CLAVATA* (*CLV*) signaling network is a key regulatory system for stem-cell proliferation in plants, which is also highly associated with crop productivity. The interaction of small secreted signaling peptides with their cognate cell-surface receptors is the hallmark of this deeply conserved network. In multiple species, it has been shown that the peptide CLV3 binds to leucine-rich repeat (LRR) receptor kinases, such as CLV1, to repress stem-cell proliferation mediated by the homeodomain protein WUSCHEL (WUS). In previous work, we reported that all conserved *CLV* genes modulate stem-cell proliferation, meristem maintenance, and fruit size in tomato. However, our more recent work has revealed important and exciting differences in *CLV* network function in tomato and related *Solanaceae* species. Specifically, we have discovered a surprising level of functional redundancy and compensation between homologous peptide and receptor genes. To dissect tomato *CLV* signaling, we characterized all predicted *CLV* components phylogenetically and used CRISPR-Cas9 technology to knock out individual and multiple genes simultaneously. Homozygous single gene knockouts of *clv1*, *clv2*, and *clv3* each showed meristem overproliferation, which led to thickened stems, increased inflorescence branching, flowers with more organs, and larger fruits. Although these phenotypes resembled corresponding mutants in other species, such as the model plant *Arabidopsis*, double mutants of *clv1 clv3* displayed a dramatic enhancement of these phenotypes, suggesting that other factors function in tomato *CLV* signaling. To explore this possibility, we performed genome-wide expression analyses on meristems of all single mutants and double-mutant combinations, which validated a previously discovered transcriptional compensation of a close homolog of *CLV3*. The *CLE9* gene has no phenotype when knocked out by itself; however, CRISPR-engineered mutations dramatically enhance *clv3* (Fig. 2). These findings are reminiscent of work in yeast and animals showing that deletion of one gene in a duplicate pair of paralogs causes transcriptional compensation of the other, forming a "responsive backup circuit" (RBC) that enhances network robustness and phenotypic stability. Our results suggest RBCs involving CLV peptides and their receptors modulate stem-cell proliferation and meristem size and raise the larger question of how prevalent RBCs are in plant genomes and to what extent they have roles in other aspects of development and evolution. Future work is aimed at searching for and characterizing RBCs in other signaling pathways.

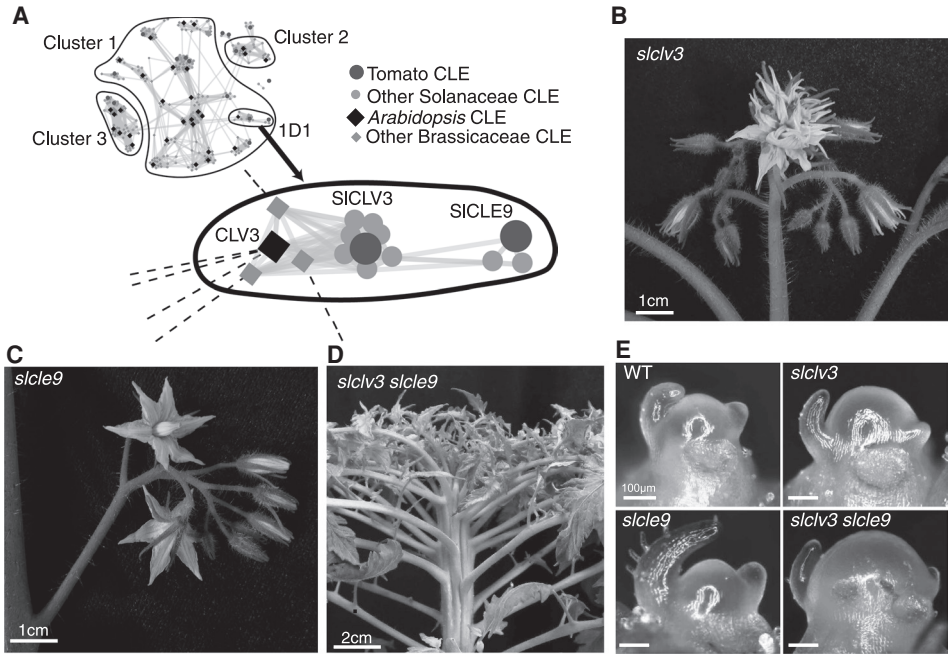


Figure 2. Redundancy and compensation in the control of stem-cell proliferation. (*A*) Clustering of CLV3 and CLE peptides showing tomato SlCLE9 is a paralog of SlCLV3. (*B*) CRISPR *slclv3* mutant. (*C*) CRISPR-generated mutants of *slcle9* resemble wild-type (WT). (*D*) *slclv3 slcle9* double mutants are dramatically enhanced. (*E*) Meristem size in each genotype. Double-mutant meristems are dramatically enlarged.

Epistatic Interactions within a Homeobox Family Controlling Meristem Maturation and Size

A. Hendelman

A key component of stem-cell homeostasis in meristems is the homeobox transcription factor gene *WUS*. Numerous studies on *WUS* and other *WUSCHEL-RELATED HOMEOBOX* (*WOX*) transcription factor genes in diverse species showed that one universal role of the *WOX* genes is controlling stem-cell proliferation with the *CLV* network. Other developmental functions include stem-cell maintenance in the root and cambium, embryo patterning, and lateral organ outgrowth. Interestingly, comparative genetics of *WOX* homologs in monocots and eudicots, and also between different eudicots, revealed cases of functional antagonism between family members. In tomato, only one of the 11 *WOX* family members has been characterized; the *COMPOUND INFLORESCENCE* gene (*S*, homolog of *Arabidopsis WOX9*) produces highly branched inflorescences when mutated because of delays in meristem maturation. To study the contribution and potential antagonistic relationships of tomato *WOX* genes on meristem maturation and proliferation in tomato, we have begun generating single and higher mutations using CRISPR. Interestingly, RNA sequencing (RNA-Seq) of *s* mutant meristems showed that tomato *WUS* is down-regulated, raising the possibility that *WUS* has a previously unknown role in maturation. Similarly, transcriptome examination of the *WOX* genes in *clv3* meristems showed that, in addition to *WUS* up-regulation, the expression of two of three *WOX3* paralogs was elevated. Not surprisingly, we found that CRISPR-generated *wus* mutants lack an enduring meristem and are epistatic to *clv3*. However, we found that at least one *WOX3* homolog has a surprising genetic interaction with *CLV3*. Given that tomato *wus* mutants fail to produce inflorescences, we are generating weak alleles using *cis*-regulatory mutagenesis (see below) to study interactions with *S*/*WOX9* in meristem maturation. Finally, tomato has three *WOX3* homologs, and higher-order mutants are being generated and studied in the context of *clv3* and other meristem maintenance mutants.

Structural Variation in Tomato Genomes and Its Role in Domestication and Crop Improvement

J. Dalrymple, X. Wang, S. Soyk, Z. Lemmon

A large component of our research involves genetic dissection of quantitative variation. A major goal of quantitative genetics is not only to identify the collection of genes that contribute to a particular complex trait but also to reveal the specific variants that define the type and strength of each allele. This knowledge is valuable for several reasons, ranging from understanding molecular mechanisms of quantitative trait loci (QTLs) to engineering customized alleles for breeding. The vast majority of single-nucleotide polymorphisms (SNPs) that have been proven to be causative for QTLs are within coding sequences and have obvious molecular consequences. Much more challenging and common is when QTLs map to *cis*-regulatory regions, where structural variants (SVs), such as insertions, duplications, deletions, inversions, and translocations, can play major roles in modifying gene expression. However, current genome-wide association study (GWAS) strategies, sequencing technologies, and computational algorithms are limited in their power to reveal the full extent and true nature of SVs, especially in noncoding regions. We are leading a project in collaboration with Joyce Van Eck, Esther van der Knaap, and Mike Schatz (Johns Hopkins University) to reveal and precisely characterize SVs in wild and domesticated tomato germplasm and study their impact on quantitative variation. We are creating many new reference genomes using single-molecule real-time (SMRT) sequencing (PacBio/Nanopore) and linked-read sequencing (10× Genomics). From these data, a compendium of SVs will be established and integrated with ongoing GWAS studies with outside collaborators. CRISPR gene knockouts and engineered SVs will be studied phenotypically, to understand impacts on growth and development, with an emphasis on genes with known or predicted roles in reproductive traits. This project will provide a foundation for dissecting SVs in other crop plants, in which such knowledge can be exploited to improve crop productivity, as described in the following sections.

Engineering Quantitative Trait Variation for Crop Improvement by Genome Editing

D. Rodriguez-Leal, Z. Lemmon

It is well known that current crop yields will not meet future food, feed, and fuel demands. Major drawbacks in plant breeding are limited genetic variation underlying quantitative traits and the time-consuming and labor-intensive evaluation of breeding germplasm required to select improved plants. Furthermore, genome sequencing and resequencing projects have revealed limitations on the genetic variation provided by nature, much of which is not useful or has subtle effects that are challenging to identify and exploit. There is, therefore, an urgent need to develop innovative approaches to accelerate crop improvement and make its outcomes more predictable. Current technologies for gene editing offer novel ways of introducing new and desired genetic variation, allowing for further improvement on plant crops by simplifying some parts of the breeding process. We have taken advantage of gene editing using CRISPR-Cas9 technology to generate new genetic variation that changes transcriptional regulation, and thus the expression, of genes associated with yield traits by targeting their upstream *cis*-regulatory sequence. We designed a genetic "drive" system that exploits heritability of CRISPR-Cas9 transgenes carrying multiple guide RNAs in "sensitized" populations to rapidly and efficiently generate dozens of novel *cis*-regulatory alleles for three genes that regulate fruit size, inflorescence architecture, and plant growth habit in tomato (Fig. 3). By segregating away the transgene in the following generation, we recovered a wide range of stabilized *cis*-regulatory alleles that provided a continuum of variation for all three traits. For one of these genes, we found that transcriptional change was a poor predictor of phenotypic effect, revealing unexplored complexity in how *cis*-regulatory variation impacts quantitative traits. Notably, plants carrying these *cis*-regulatory alleles showed increased productivity when compared with other plants carrying high-order combinations of mutant alleles from different genes previously described to improve yield. These results constitute the first example of using genome-editing technologies to engineer new and desirable quantitative trait variation with immediate benefits for breeding, providing a powerful and widely applicable approach for boosting crop productivity in many crops. Our current

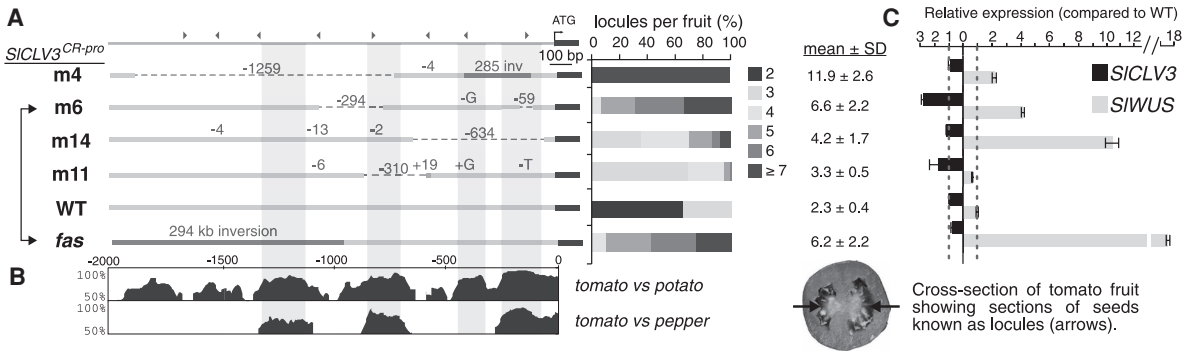


Figure 3. CRISPR-generated promoter QTL alleles for the tomato *CLV3* (*SlCLV3*) stem-cell gene. (*A*) A subset of six alleles is shown reflecting a continuum of variation for fruit locule number. Allele "m6" recreates the effect of the natural "fasciated" (fas) QTL that increased fruit size during domestication. (*B*) mVISTA plots comparing three Solanaceae species showing conserved noncoding sequences (CNSs) aligned with the promoter alleles. (*C*) Expression (qRT-PCR) of *SlCLV3* and its downstream target *SlWUS*. More information available in Rodriguez-Leal *et al.* (2017).

work is aimed at determining whether this strategy can help bypass the time and effort breeders expend trying to adapt beneficial allelic variants to diverse breeding germplasm.

Engineering Crop Domestication

J. Dalrymple, Z. Lemmon

Creating new crops from wild plants with agricultural potential could greatly contribute to future food security efforts. Yet, developing breeding programs for orphan crops is expensive in terms of time, labor, and financial investment. An exciting new prospect is using genome-editing technologies to perform targeted trait improvement, bypassing the disadvantages of a traditional breeding pipeline. Comparative biology between a potential crop and related species can also reveal the genes and pathways critical to agronomic success. For example, recurring themes in the domestication and improvement of many crops are modification of the timing and style of flowering and the control of meristem size influencing fruit and seed production.

The tomato *Solanaceae* cousin goldenberry (*Physalis peruviana*) is an emerging berry crop with subtly sweet and tangy, highly nutritious fruits that are enclosed in an inflated calyx or "husk." However, goldenberry has never been domesticated, and thus produces relatively small fruits on extremely large plant architectures (Fig. 4). We are collaborating with J. Van Eck at the Boyce Thompson Institute to show the potential of rapidly and efficiently improving *Physalis* by genome editing, facilitated by foundational genomic and transformation resources we have generated, including genome transcriptome sequencing. These

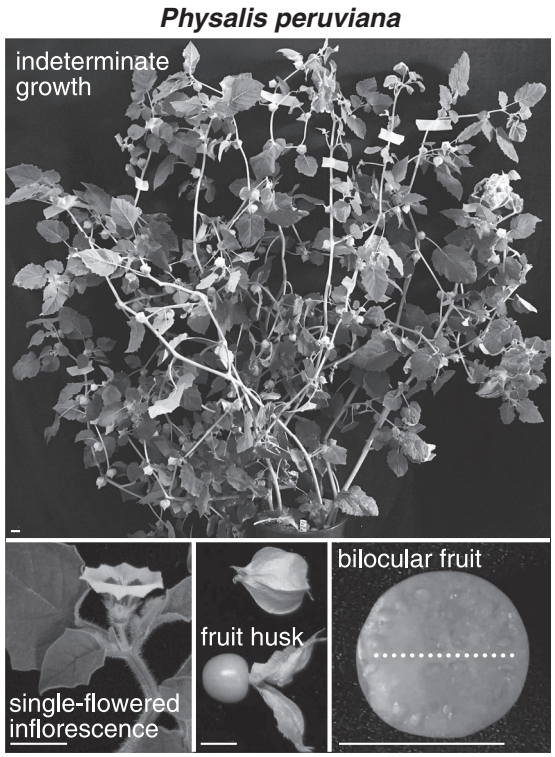


Figure 4. The Solanaceae species *Physalis peruviana* chosen for engineering domestication. *P. peruviana* is a large weedy plant that produces single-flower inflorescences and bears tiny (1-gm) bilocular fruits with an outer husk. Scale bars, 1 cm.

data were used for de novo assemblies, and orthologs of key tomato genes that impact agricultural traits were identified. CRISPR constructs were designed to target *Physalis* orthologs of multiple agronomic genes and traits, including determinate growth, increased fruit size, and modified fruit color. We have successfully generated mutations in several genes, and we are now developing constructs to engineer QTLs using our CRISPR promoter technology. A notable achievement has been the generation of heritable genome-edited null alleles of the *Physalis CLV1* ortholog, which resulted in a 20% increase in fruit size. Our work also provides a valuable genomic resource in a relatively unstudied clade of the *Solanaceae*, which has already been leveraged to explore the evolution of Solanaceous signaling peptide and MADS-box transcription factor gene families.

PUBLICATIONS

Rodriguez-Leal D, Lemmon Z, Man J, Bartlett ME, Lippman ZB. 2017. Engineering quantitative trait variation for crop improvement by genome editing. *Cell* **171:** 470–480.

Soyk S, Lemmon Z, Oved M, Fisher J, Liberatore KL, Park SJ, Goren A, Jiang K, Ramos A, van der Knaap E, et al. 2017. Bypassing negative epistasis on yield in tomato imposed by a domestication gene. *Cell* **169:** 1142–1155.

EPIGENETIC INHERITANCE IN PLANTS AND FISSION YEAST

R. Martienssen B. Berube, S. Bhattacharjee, F. de Sousa Borges, E. Ernst, M. Gutbrod, R. Herridge, H.S. Kim, S-C. Lee, J-S. Parent, U. Ramu, M. Regulski, J. Ren, B. Roche, A. Schorn, A. Shimada, J. Simorowski

Plants and fission yeast provide excellent models for epigenetic mechanisms of transposon regulation, heterochromatic silencing, and gene imprinting, important for both plant breeding and human health. We are investigating the role of RNA interference (RNAi) in heterochromatic silencing in the fission yeast *Schizosaccharomyces pombe* and the model plant *Arabidopsis thaliana*, as well as continuing our work on *Zea mays*. In fission yeast, we have found that RNAi promotes DNA replication and repair, as well as histone modification. In quiescence, RNAi becomes essential because it is required for release of RNA polymerase from ribosomal DNA repeats. Genetic screens have revealed dependency on transcription, heterochromatin, and chromosome segregation and provide insight into *DICER1* mutations in cancer. We have found that long terminal repeat (LTR) retrotransposons in the mouse are controlled by fragments of transfer RNA (tRNA) that match the highly conserved primer binding site (PBS), where retrotransposons are primed for reverse transcription by intact tRNA, revealing an ancient link between RNAi, translation, and retrotransposition. Remarkably, plant retrotransposons are similarly targeted by a microRNA (miRNA) in pollen. This miRNA is required for the triploid dosage response, a barrier to hybridization first discovered by Albert Blakeslee at the Cold Spring Harbor Station for Experimental Evolution 100 years ago. We continue to develop duckweeds for biofuel by sequencing the genome and developing an efficient transformation system in the clonally propagated aquatic macrophyte *Lemna minor*. This year we said goodbye to postdocs Rowan Herridge and Filipe Borges, who left for positions at the University of Central Otago in New Zealand and INRA Versailles in Paris, respectively. Dr. Chris Winefield from Lincoln University visited on study leave and returned to his research in grapevine genomics in New Zealand.

Transcriptional Reprogramming in Cellular Quiescence

B. Roche, R.A. Martienssen [in collaboration with B. Arcangioli, Institut Pasteur, Paris]

Most cells in nature are not actively dividing, yet are able to return to the cell cycle given the appropriate environmental signals. There is now ample evidence that quiescent G_0 cells are not shut down but still metabolically and transcriptionally active. Quiescent cells must maintain a basal transcriptional capacity to maintain transcripts and proteins necessary for survival. This implies a tight control over RNA polymerases: RNA Pol II for messenger RNA (mRNA) transcription during G_0, but especially RNA Pol I and RNA Pol III to maintain an appropriate level of structural RNAs, raising the possibility that specific transcriptional control mechanisms evolved in quiescent cells. In accordance with this, we recently discovered that RNAi is necessary to control RNA Pol I transcription during G_0. The nucleolus is a distinct compartment of the nucleus responsible for ribosome biogenesis. Misregulation of nucleolar functions and the cellular translation machinery has been associated with disease—in particular, many types of cancer. Indeed, many tumor suppressors (p53, Rb, PTEN, PICT1, BRCA1) and proto-oncogenes (MYC, NPM) play a direct role in the nucleolus and interact with the RNA Pol I transcription machinery and the nucleolar stress response. We have identified Dicer and the RNAi pathway as having essential roles in the nucleolus of quiescent *S. pombe* cells, distinct from pericentromeric silencing, by controlling RNA Pol I release. We propose that this novel function is evolutionarily conserved and may contribute to the tumorigenic predisposition of *DICER1* mutations in mammals.

The Conserved RNA-Binding Cyclophilin, Rct1, Regulates Small RNA Biogenesis and Splicing Independent of Heterochromatin Assembly

E. Ernst, H.S. Kim, R.A. Martienssen [in collaboration with A.Y. Chang, Boston Biomedical, Cambridge, MA; S.E. Castel, New York Genome Center]

RNAi factors and their catalytic activities are essential for heterochromatin assembly in *S. pombe*. This has led to the idea that small interfering RNAs (siRNAs) can promote H3K9 methylation by recruiting the *cryptic loci regulator* complex (CLRC), also known as *recombination in K* complex (RIKC), to the nucleation site. The conserved RNA-binding protein Rct1 (AtCyp59/SIG-7) interacts with splicing factors and RNA Pol II. Here we show that Rct1 promotes processing of pericentromeric transcripts into siRNAs via the RNA recognition motif. Surprisingly, loss of siRNA in *rct1* mutants has no effect on H3K9 dimethylation or trimethylation, resembling other splicing mutants, suggesting that posttranscriptional gene silencing per se is not required to maintain heterochromatin. Splicing of the Argonaute gene is also defective in *rct1* mutants and contributes to loss of silencing but not to loss of siRNA. Our results suggest that Rct1 guides transcripts to the RNAi machinery by promoting splicing of elongating noncoding transcripts.

LTR-Retrotransposon Control by tRNA-Derived Small RNAs

A.J. Schorn, M.J. Gutbrod, R.A. Martienssen [in collaboration with C. LeBlanc, Yale University]

Transposon reactivation is an inherent danger in cells that lose epigenetic silencing during developmental reprogramming. In the mouse, LTR retrotransposons, or endogenous retroviruses (ERV), account for most novel transposon insertions and are expressed in the absence of histone H3K9 trimethylation in preimplantation embryonic and trophoblast stem cells. tRNA fragments (tRFs) are a class of small, regulatory RNAs with diverse functions. 3′-derived tRFs perfectly match LTR retroelements, which use the 3′ end of full-length tRNAs to prime reverse transcription. We have found abundant 18-nt 3′ tRF in these cells and ubiquitously expressed 22-nt 3′ tRFs that include the 3′ terminal CCA of mature tRNAs and target the specific tRNA PBS essential for ERV reverse transcription. We show that the two most active ERV families in the mouse, IAP and MusD/ETn, are major targets and strongly inhibited by tRFs in retrotransposition assays. Twenty-two-nucleotide tRFs posttranscriptionally silence coding-competent ERVs, whereas 18-nt tRFs specifically interfere with reverse transcription and retrotransposon mobility. The highly conserved tRNA PBS in LTR retroelements is a unique target for 3′ tRFs to recognize and block abundant but otherwise highly diverse LTR retrotransposons. 3′ tRFs are processed from full-length tRNAs under so far unknown conditions and potentially protect many cell types (Fig. 1). tRFs are often bound by small RNA binding proteins in the Argonaute family and appear to be an ancient link among RNAi, transposons, and genome stability.

Transposon-Derived Small RNAs Triggered by miR845 Mediate Genome Dosage Response in *Arabidopsis*

F. de Sousa Borges, J-S. Parent, R.A. Martienssen [in collaboration with F. van Ex, Bayer Crop Science, Belgium; C. Köhler, Linnean Center of Plant Biology, Swedish University of Agricultural Sciences, Uppsala]

Chromosome dosage has substantial effects on reproductive isolation and speciation in both plants and animals, but the underlying mechanisms are largely obscure. Transposable elements in animals can regulate hybridity through maternal small RNA, whereas small RNAs in plants have been postulated to regulate dosage response via neighboring imprinted genes. We have found that a highly conserved miRNA in plants, miR845, targets the $tRNA^{Met}$ PBS of LTR retrotransposons in *Arabidopsis* pollen, mimicking the role of tRNA fragments in mammals. miR845b triggers the accumulation of 21- to 22-nt small RNAs in a dose-dependent fashion via a noncanonical pathway requiring RNA polymerase IV (Pol IV). We show that these epigenetically activated small interfering RNAs (easiRNAs) mediate hybridization barriers between diploid seed parents and tetraploid pollen parents (the "triploid block"), and natural variation for miR845 may account for "endosperm balance," allowing the formation of triploid seeds. Thus, targeting of the PBS with small RNA is a common mechanism for transposon control in mammals and plants and provides a uniquely sensitive means to monitor chromosome dosage and imprinting in the developing seed.

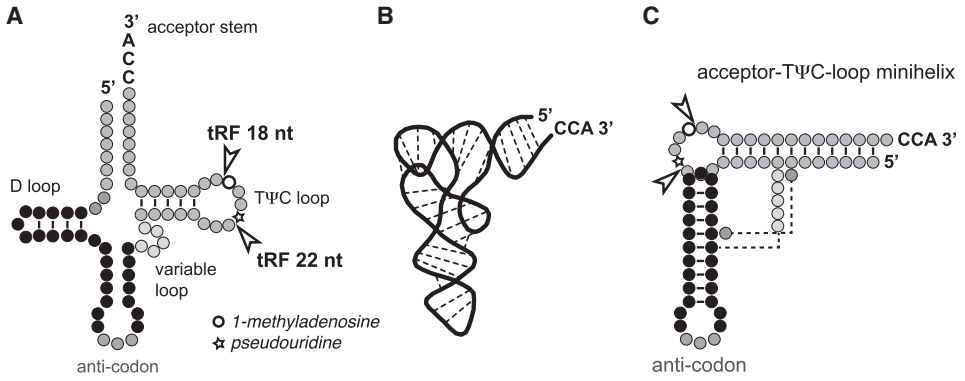


Figure 1. Transfer RNA (tRNA) structure and domains. Three alternative representations of the same tRNA, here of tRNA Lys-AAA with the anticodon-sequence UUU, illustrate its functional domains and tertiary structure. (*A*) The typical cloverleaf diagram provides an overview of the major domains and loops. The acceptor stem has the post-transcriptional CCA-tag at its 3′ end that will be covalently linked with the amino acid before translation. (*B*) The three-dimensional (3D) L-shaped structure of tRNA is composed of two double-stranded helices. (*C*) A flattened minihelix representation of the 3D structure, showing the anticodon-D-loop (black/gray) that serves proofreading functions during translation at the ribosome and the upper minihelix (light gray) consisting of the acceptor-stem and the TYC-loop. Modified bases adjacent to the processing sites for 3′ tRNA derived fragments (arrows) are shown.

Live-Cell Analysis of DNA Methylation during Sexual Reproduction in *Arabidopsis* Reveals Context- and Sex-Specific Dynamics Controlled by Noncanonical RdDM

A.J. Schorn, R.A. Martienssen [in collaboration with D. Grimanelli and M. Ingouff, Institut de Recherche pour le Développement, Montpellier; F. Berger, Gregor Mendel Institute, Vienna]

Cytosine methylation is a key epigenetic mark in many organisms, important for both transcriptional control and genome integrity. Although relatively stable during somatic growth, DNA methylation is reprogrammed genome-wide during mammalian reproduction. Reprogramming is essential for zygotic totipotency and to prevent transgenerational inheritance of epimutations. However, the extent of DNA methylation reprogramming in plants remains unclear. Our collaborators have developed sensors reporting with single-cell resolution CG and non-CG methylation in *Arabidopsis*. Live imaging during reproduction revealed distinct and sex-specific dynamics for both contexts. We found that CHH methylation in the egg cell depends on *DOMAINS REARRANGED METHYLASE 2* (DRM2) and RNA polymerase V (Pol V), two main actors of RNA-directed DNA methylation, but does not depend on Pol IV. We have also been able to visualize methylation in living mouse embryonic stem cells, opening the door to live imaging of genome reprogramming in vitro. Our sensors provide insight into global DNA methylation dynamics at the single-cell level with high temporal resolution and offer a powerful tool to track CG and non-CG methylation both during development and in response to environmental cues in all organisms with methylated DNA.

Heterochromatic Silencing Mechanisms and Retrotransposon Controlled by DDM1 and 21- to 22-nt easiRNAs

S-C. Lee, E. Ernst, R. Herridge, J-S. Parent, R.A. Martienssen [in collaboration with D. Jackson, Cold Spring Harbor Laboratory; D. Grimanelli, Institut de Recherche pour le Développement, Montpellier]

Plant heterochromatin is mainly composed of transposons and repeats, which contain epigenetic signatures, including DNA methylation and histone modifications such as H3K9me2 and H3K27me1. H3.1 and H3.3 are two major H3 variants dominantly deposited on heterochromatin and euchromatin, respectively. H3.1 is specifically modified with monomethylation of H3K27 by two histone methyltransferases, ATXR5 and ATXR6. We have found that the SNF2 chromatin remodeler DDM1 is involved in the deposition of histone variants in *Arabidopsis*. H3.1 incorporation and H3K27me1 modification were markedly reduced in *ddm1*, whereas H3.3 accumulation was enhanced in heterochromatin. The localization patterns of H3.1 and H3.3 in *ddm1* were reminiscent of *fas2*, which

encodes the CAF-1 chaperone regulating H3.1 deposition. H3.1 is incorporated into chromatin during S phase and evicted after cell differentiation. DDM1 is incorporated into chromatin during G_1 and evicted during mitosis. *ddm1* mutants have massive transcriptional activation of transposons, resulting in 21- to 22-nt easiRNAs produced by RDR6. We performed polysomal RNA sequencing (RNA-seq) and virus-like-particle (VLP) DNA-seq to investigate functions of 21- to 22-nt easiRNAs in translation and reverse transcription. A handful of *ATHILA* family elements were differentially regulated between *ddm1* and *ddm1rdr6* in translation and reverse transcription, but the majority of transposable elements remained unaffected. Functional LTR retrotransposons that produce full-length VLP DNA were mostly from *ATCOPIA* families, including *EVADE*. Unexpectedly, DDM1 was involved in de novo silencing of activated *EVADE* elements in female gametophytes, along with AGO9 and RDR6.

Natural Variation and Dosage of the HEI10 Meiotic E3 Ligase Control *Arabidopsis* Crossover Recombination

R.A. Martienssen [in collaboration with C.J. Underwood, KeyGene, Wageningen, the Netherlands; I.R. Henderson, University of Cambridge; F.C. Franklin, University of Birmingham]

During meiosis, homologous chromosomes undergo crossover recombination, which creates genetic diversity and balances homolog segregation. Despite these critical functions, crossover frequency varies extensively within and between species. Although natural crossover recombination modifier loci have been detected in plants, causal genes have remained elusive. Using two *A. thaliana* accessions, we identified two major recombination quantitative trait loci (rQTL) that explain 56.9% of crossover variation in F_2 populations. We mapped rQTL1 to semidominant polymorphisms in *HEI10*, which encodes a conserved ubiquitin E3 ligase that regulates crossovers. Null *hei10* mutants are haploinsufficient, and using genome-wide mapping and immunocytology, we showed that transformation of additional HEI10 copies is sufficient to more than double euchromatic crossovers. However, heterochromatic centromeres remained recombination suppressed. The strongest HEI10-mediated crossover increases occur in subtelomeric euchromatin, which is reminiscent of sex differences in *Arabidopsis* recombination. Our work reveals that HEI10 naturally limits *Arabidopsis* crossovers and has the potential to influence the response to selection.

Nucleosomes and DNA Methylation Shape Meiotic DSB Frequency in *A. thaliana* Transposons and Gene Regulatory Regions

C.J. Underwood, R.A. Martienssen [in collaboration with I.R. Henderson, University of Cambridge]

Meiotic recombination initiates from DNA double-strand breaks (DSBs) generated by SPO11 topoisomerase-like complexes. Meiotic DSB frequency varies extensively along eukaryotic chromosomes, with hotspots controlled by chromatin and DNA sequence. To map meiotic DSBs throughout a plant genome, we purified and sequenced *A. thaliana* SPO11-1-oligonucleotides. SPO11-1-oligos are elevated in gene promoters, terminators, and introns, which is driven by AT-sequence richness that excludes nucleosomes and allows SPO11-1 access. A positive relationship was observed between SPO11-1-oligos and crossovers genome wide, although fine-scale correlations were weaker. This may reflect the influence of interhomolog polymorphism on crossover formation, downstream from DSB formation. Although H3K4me3 is enriched in proximity to SPO11-1-oligo hotspots at gene 5′ ends, H3K4me3 levels do not correlate with DSBs. Unexpectedly, we found high SPO11-1-oligo levels in nucleosome-depleted Helitron/Pogo/Tc1/Mariner DNA transposons, whereas retrotransposons were coldspots. High SPO11-1-oligo transposons are enriched within gene regulatory regions and in proximity to immunity genes, suggesting a role as recombination enhancers. Epigenetic activation of meiotic DSBs in proximity to centromeres and transposons occurred in *met1* CG DNA methylation mutants, coincident with reduced nucleosome occupancy, gain of transcription, and H3K4me3. However, crossover levels were suppressed in these regions.

Epigenetic Activation of Meiotic Recombination near Centromeres via Loss of H3K9me2 and Non-CG DNA Methylation

C.J. Underwood, F. de Sousa Borges, J. Simorowski, E. Ernst, R.A. Martienssen [in collaboration with Y. Jacob, Yale University; I.R. Henderson, University of Cambridge]

Eukaryotic centromeres contain the kinetochore, which connects chromosomes to the spindle to allow

segregation. During meiosis, centromeres are suppressed for interhomolog crossover, as recombination in these regions can cause chromosome missegregation and aneuploidy. Plant centromeres are surrounded by transposon-dense pericentromeric heterochromatin that is epigenetically silenced by H3K9me2 and DNA methylation in CG and non-CG sequence contexts. However, the role of these chromatin modifications in control of meiotic recombination in the pericentromeres is not fully understood. We have shown that disruption of *A. thaliana* H3K9me2 and non-CG DNA methylation pathways, via mutation of the H3K9 methyltransferase genes *KYP/SUVH4 SUVH5 SUVH6*, or the CHG (H = A, T, or C) DNA methyltransferase gene *CMT3*, increases meiotic recombination in proximity to the centromeres. We observed that H3K9me2 and non-CG DNA methylation pathway mutants show increased pericentromeric crossovers in hybrid and inbred backgrounds and involve both the interfering and noninterfering crossover repair pathways. We also showed that meiotic DSBs increase in H3K9me2/non-CG mutants within the pericentromeres, via purification and sequencing of SPO11-1-oligonucleotides. Therefore, H3K9me2 and non-CG DNA methylation exert a repressive effect on both meiotic DSB and crossover formation in plant pericentromeric heterochromatin. Our results may account for strong selection of enhancer trap Dissociation (Ds) transposons into the *CMT3* gene by recombination with proximal transposon launchpads.

Genome-Wide Analysis of *Arabidopsis* and Maize Replication Timing Programs

C. LeBlanc, R. Martienssen [in collaboration with G.C. Allen, L. Hanley-Bowdoin, and W.F. Thompson, North Carolina State University, Raleigh; M.W. Vaughn, University of Texas, Austin; H. Bass, Florida State University]

Eukaryotes use a temporally regulated process, known as the replication timing program, to ensure that their genomes are fully and accurately duplicated during S phase. Although replication timing programs have been described for yeast and animal systems, much less is known about the temporal regulation of plant DNA replication. We used the thymidine analog, 5-ethynyl-2′-deoxyuridine, in combination with flow sorting and Repli-Seq to describe, at high-resolution, the genome-wide replication timing program for *Arabidopsis* (*A. thaliana*) suspension cells. We identified genomic regions that replicate during early, mid, and late S phase and correlated these regions with chromatin state, accessibility, and long-distance interaction. *Arabidopsis* chromosome arms tend to replicate early, whereas pericentromeric regions replicate late. Early and mid-replicating regions are gene-rich and predominantly euchromatic, whereas late regions are rich in transposable elements and primarily heterochromatic. The distribution of chromatin states is complex, with each replication time corresponding to a mixture of states. Intact root tips of maize include several different cell lineages, and we have determined whole-genome replication timing profiles from cells in early, mid, and late S phase of the mitotic cell cycle, which each constitute between 20% and 24% of the genome, whereas ~32% of the genome shows replication activity in two different time windows. The centromere core replicates in mid S, before more heavily compacted classical heterochromatin, including pericentromeres and knobs, which replicate during late S phase.

Highly Contiguous Duckweed Genomes from Long Read Sequencing

E. Ernst, R.A. Martienssen [in collaboration with Eric Lam, Rutgers University; J. Shanklin, Brookhaven National Laboratory]

Lemnaceae are the world's smallest aquatic flowering plants, and although they are true monocotyledonous angiosperms, they have a highly reduced morphology comprising growing fronds, resting fronds, simple roots, and a meristem-like "pocket." *Lemnaceae* in optimal conditions have an exponential growth rate that can double the number of fronds in 24 hours and produce 64 g of biomass per gram starting weight in a week, which is far beyond the fastest growing corn rates (2.3 g/g/week), and unencumbered by secondary products such as lignin. Our goal is to divert a substantial portion of accumulated carbon from starch to oil metabolism in *Lemna*, using resting fronds and primitive roots as the storage tissue. Clonal propagation, limited seed set, and variable chromosomal number are shared with sugarcane and *Miscanthus gigantea*, and many of the design principles and technologies we develop will have applications in other energy crops. The maturation of long single molecule sequencing technologies has enabled non–model

organism genomes to be assembled with contiguity better than that of existing reference genomes at vastly lower cost. We have completed the first long read assemblies of the *Lemna gibba* 7742 and *Lemna minor* 8627 genomes, along with updated gene annotations informed by full-length cDNA sequences from the Oxford Nanopore MinION. Using these improved assemblies, we have examined variation in heterozygosity, repeat content, and DNA methylation among the *Lemnaceae*. Genomics has greatly impacted breeding in domesticated row crops, through genome and marker assisted selection, as well as genes underlying key traits. But crops with only minimal domestication, such as oil palm and duckweed, present challenges to this approach, especially when they are propagated asexually as clones. Although clones have the advantage of potentially fixing hybrid vigor, lack of germline passage, in which epigenetic reprogramming occurs, can lead to epigenetic variation.

PUBLICATIONS

Chang AY, Castel SE, Ernst E, Kim HS, Martienssen RA 2017. The conserved RNA binding cyclophilin, Rct1, regulates small RNA biogenesis and splicing independent of heterochromatin assembly. *Cell Rep* **19:** 2477–2489.

Chomet P, Martienssen R. 2017. Barbara McClintock's final years as nobelist and mentor: A memoir. *Cell* **170:** 1049–1054.

He H, Li Y, Dong Q, Chang AY, Gao F, Chi Z, Su M, Zhang F, Ban H, Martienssen R, et al. 2017. Coordinated regulation of heterochromatin inheritance by Dpb3–Dpb4 complex. *Proc Natl Acad Sci* **114:** 12524–12529.

Ingouff M, Selles B, Michaud C, Vu TM, Berger F, Schorn AJ, Autran D, Van Durme M, Nowack MK, Martienssen RA, et al. 2017. Live-cell analysis of DNA methylation during sexual reproduction in *Arabidopsis* reveals context and sex-specific dynamics controlled by noncanonical RdDM. *Genes Dev* **31:** 72–83.

Roche B, Arcangioli B, Martienssen R. 2017. Transcriptional reprogramming in cellular quiescence. *RNA Biol* **14:** 843–853.

Roche B, Arcangioli B, Martienssen R. 2017. New roles for Dicer in the nucleolus and its relevance to cancer. *Cell Cycle* **16:** 1643–1653.

Schorn AJ, Gutbrod MJ, LeBlanc C, Martienssen R. 2017. LTR-retrotransposon control by tRNA-derived small RNAs. *Cell* **170:** 61–71.e11.

Underwood CJ, Henderson IR, Martienssen RA. 2017. Genetic and epigenetic variation of transposable elements in *Arabidopsis*. *Curr Opin Plant Biol* **36:** 135–141.

Wear EE, Song J, Zynda GJ, LeBlanc C, Lee TJ, Mickelson-Young L, Concia L, Mulvaney P, Szymanski ES, Allen GC, et al. 2017. Genomic analysis of the DNA replication timing program during mitotic S phase in maize (*Zea mays*) root tips. *Plant Cell* **29:** 2126–2149.

Ziolkowski PA, Underwood CJ, Lambing C, Martinez-Garcia M, Lawrence EJ, Ziolkowska L, Griffin C, Choi K, Franklin FC, Martienssen RA, et al. 2017. Natural variation and dosage of the HEI10 meiotic E3 ligase control *Arabidopsis* crossover recombination. *Genes Dev* **31:** 306–317.

In Press

Borges F, Parent JS, van Ex F, Wolff P, Martínez G, Köhler C, Martienssen RA. 2018. Transposon-derived small RNAs triggered by miR845 mediate genome dosage response in *Arabidopsis*. *Nature Genet* **50:** 186–192.

Choi K, Zhao X, Tock AJ, Lambing C, Underwood CJ, Hardcastle TJ, Serra H, Kim J, Cho HS, Kim J, et al. 2018. Nucleosomes and DNA methylation shape meiotic DSB frequency in *Arabidopsis thaliana* transposons and gene regulatory regions. *Genome Res* **28:** 532–546.

Concia L, Brooks AM, Wheeler E, Zynda GJ, Wear EE, LeBlanc C, Song J, Lee TJ, Pascuzzi PE, Martienssen RA, et al. 2018. Genome-wide analysis of the *Arabidopsis* replication timing program. *Plant Physiol* **176:** 2166–2185.

Schorn AJ, Martienssen R. 2018. Tie-break: Host and retrotransposons play tRNA. *Trends Cell Biol* doi: 10.1016/j.tcb.2018.05.006.

Underwood CJ, Choi K, Lambing C, Zhao X, Serra H, Borges F, Simorowski J, Ernst E, Jacob Y, Henderson IR, et al. 2018. Epigenetic activation of meiotic recombination near *Arabidopsis thaliana* centromeres via loss of H3K9me2 and non-CG DNA methylation. *Genome Res* **28:** 519–531.

MOLECULAR SIGNALING EVENTS UNDERLYING ENVIRONMENTAL CONTROL OF PLANT GROWTH

U. Pedmale L.N. Lindbäck A. Matheus K. O'Neill S. Sankaranarayanan K. Schwartz N. Shanmugarajah O. Spassibojko

A fundamental question in biology, which remains unanswered, is how the environment of the organism regulates its growth and development. Both plants and animals interact with their environment; however, plants grow postembryonically as they are incapable of moving around. Unlike animals, plants do not have specific organs that see or hear various stimuli, yet plants are sensitive to their surrounding environment and modify their growth according to various external and internal signals. Plants regularly face variability in growth conditions—temperature, light quality and quantity, herbivores, pathogens, water availability, etc. Plants respond to these biotic and abiotic factors and survive substantial fluctuations in their environment. Plants also must balance the range of potential threats and benefits confronting them and should make appropriate decisions on resource allocation. Remarkably given their lack of a brain, plants can successfully integrate various cues and make appropriate decisions about growth. Such adaptability is essential to the sessile nature of the plants. In some adaptive responses—for example, when the plants have to cope with climate change and increased competition for light—there is a decrease in productivity (yield, biomass) as the plant reallocates resources to adapt better.

The goal of our laboratory is to identify the mechanisms behind how a plant perceives and successfully adapts to its environment. We also aim to understand how a plant must integrate intrinsic and extrinsic cues and "decide" how best to respond to environmental cues. Understanding how plants deal with and respond to a multitude of environmental signals could help to develop crops that cope with unfavorable growth conditions without significant changes in yield. Our laboratory primarily studies the effect of environmental light on plant growth and development. Light is among the most relevant environmental signals because it not only drives photosynthesis, but also provides critical information about the local growth environment as well as seasonal time. Light is perceived by a complex array of photoreceptors, which include phytochromes (PHYA-E), cryptochromes (CRY1-2), phototropins (PHOT1-2), zietlupe family (FKF1, LKP2, and ZTL), and UVR8. Plants have developed various adaptive responses to interpret and utilize light directionality, quantity, and quality. In vegetational shading, when plants are under the shade of another plant, they perceive a decrease in the ratio of red to far-red light (R:FR) because of absorption of red light by chlorophyll and reflection of far-red light by the neighboring foliage. Simultaneously, there is also decrease in blue light and the available photosynthetically active radiation (PAR).

We focus on blue light–absorbing CRYs. Apart from being an excellent genetic and molecular tool to tease out the complexities of growth and adaptation, there are still many open questions about the molecular function of CRYs in plants. Understanding the role of CRYs is not only appealing for agriculture, but it also has an impact on human health, which could make this field appealing to diverse funding agencies. CRYs regulate growth and development and provide circadian entrainment to both plants and animals. In metazoans, disruption of CRY activity is linked to cancer, altered behavior, magnetoreception, and metabolism. Therefore, understanding CRY function in plants is not only important for its role in plant growth, but also for its relevance to human health.

How Does the Above-Ground Shoot Control Below-Ground Root Growth?

During shading (Fig. 1), many aerial organs elongate rapidly, whereas root growth is reduced with the delay in the emergence of lateral roots. Roots not only serve as a mechanical anchor, but also play a vital role in

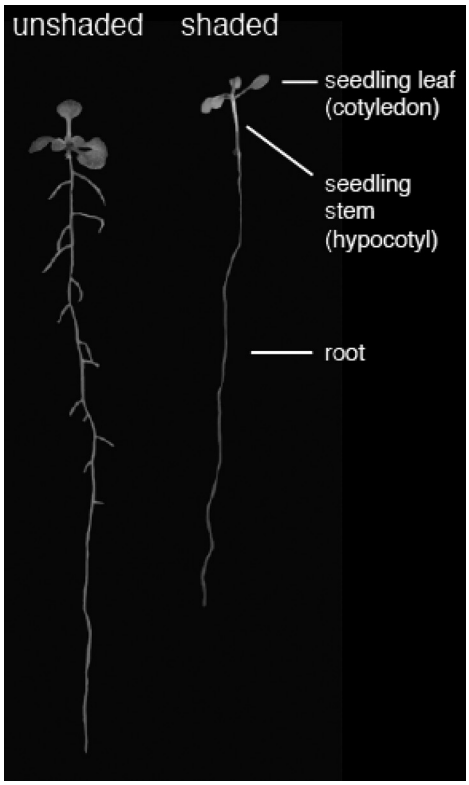


Figure 1. Light quality and quantity control of plant growth. *Arabidopsis* seedlings of the same age and genotype, but grown in separate light conditions, shaded and unshaded, have different body plan. Under shaded conditions, the seedlings rapidly elongate their stem and grow taller to outcompete their neighbors to maximize light capture at the cost of root and leaf growth.

the well-being of the entire plant. Therefore, a robust and well developed root system is required for healthy plant growth. As one can imagine, there is a negative cycle occurring during shading; shoot-perceived shade leads to reduced root growth, which in turn is unable to support the shoot, leading to unproductive plants. However, this phenomenon is an excellent model to understand growth at a systems level because of the different growth phenotypes observed in the various organs of the same plant. Additionally, the ability to explore the nature of the interorgan and long-distance communication that is used to signal when a distant organ is exposed to an adverse environment is not well understood. Unfortunately, and surprisingly, not much is known about the mechanisms that underlie reduced root growth seen during shading.

During shading, newly synthesized auxin hormone in the cotyledons is required for hypocotyl growth. First, we measured free auxin levels in the dissected root and shoot of seedlings exposed to low R:FR. As expected, we saw an increase in free auxin in the shoot and, interestingly, a 10-fold accumulation in the root. However, we failed to see the activity of various auxin reporters in the root compared to the hypocotyl. High auxin levels are known to stimulate lateral root growth, but intriguingly we did not detect growth of lateral roots; however, lateral root primordia were present. These results indicate that although auxin accumulation occurred, auxin signaling or perception was blocked in the root during shading. We also performed broad time course transcriptomic analysis from excised cotyledons, hypocotyl, and roots from *Arabidopsis* seedlings exposed to shade and nonshading conditions. Our transcriptomic analysis detected induction of auxin-responsive genes in the hypocotyl, but not in the root. Interestingly, genes induced during abiotic and biotic stress responses were significantly up-regulated in the root. To further test whether the roots in shaded plants activated stress responses to slow down their growth, we also employed various reporters that are activated during stress or defense against pathogens. Therefore, our results suggest that stress responses in the root are likely inhibiting its growth. Furthermore, it is known that the plant can either grow or defend, and it is probable that in shaded plants, by turning on the defense or stress response in the roots, resources are diverted to the shoot organs, which then compete for light. However, the molecular determinant that switches a plant between these two states of defense/stress and growth is unknown and is of immense interest to biologists. We are taking several approaches to identify this molecular switch.

Molecular Determinants of CRY2 Protein Signaling and Stability

CRYs were first identified in plants and then discovered in animals. CRY2 protein accumulates in the dark and in vegetational shade and is readily degraded by the 26S proteasome under high intensities of blue light. Therefore, it is obvious that the CRY2 protein level and activity are tightly regulated by light in order to ensure proper signaling and response. However, the signaling events from the photoactivated CRYs to growth and development programs are not known. In animals and plants, CRY protein levels and activity are tightly modulated to influence signaling outcome.

Therefore, to elucidate CRY signaling pathway, our laboratory has purified CRY2-containing protein complexes from *Arabidopsis thaliana* seedlings exposed to low intensity of blue light, which is typically encountered under shading. We identified CRY2-associated proteins by tandem affinity purification and mass spectrometry. Interestingly, the orthologs of some of the CRY2-associated proteins were also present in CRY protein complexes purified from human cells. This indicates that there could be a similar signaling mechanism in these two different evolutionary lineages.

One of the highly enriched proteins in the CRY2-associated protein complex was a deubiquitinase (DUB), which removes the ubiquitin protein covalently bound to a target protein. We found out that CRY2 and these deubiquitinases can interact directly in the nucleus of the cell. We hypothesize that CRY2 is ubiquitinated continuously, but in certain situations, it recruits deubiquitinases to protect itself from proteasomal degradation such that downstream signaling can proceed. Next, we tested CRY2 protein levels in the deubiquitinase mutants and in plants in which they are overexpressed. Surprisingly, we found that CRY2 protein levels were very high in the deubiquitinase mutant and lower when overexpressed (Fig. 2). This matches with the physiological response exhibited by the seedling stem length of these genetic backgrounds. The deubiquitinase mutant seedling had a short hypocotyl when compared to the wild-type and the *cry2* mutant, and the overexpression line had a longer hypocotyl similar to *cry2* (Fig. 2). This observation indicates that the deubiquitinase-CRY2 protein complex likely modifies a protein that affects CRY2 protein levels like a E3 ubiquitin ligase. Efforts are under way to identify ubiquitylated residues in the CRY2 protein and to evaluate the effect of catalytically dead deubiquitinases in plants. Using seedlings expressing catalytically dead deubiquitinase, we have identified several candidate E3 ligases that may affect CRY2 protein levels. Unlike animals, substrates for the large number (approximately 50) of plant deubiquitinases remain unidentified, except for histones. Therefore, we are presented with a unique opportunity to study CRY2 as one of the novel substrates of plant deubiquitinases and its role in plant growth and development. In parallel, to identify the E3 ubiquitin ligase responsible for targeting plant CRY2 for degradation, we are undertaking a forward genetic screen. In conclusion, our findings will provide novel insight into the regulation of CRYs by reversible ubiquitination as well as the role of DUBs in plant development, which is largely unknown so far.

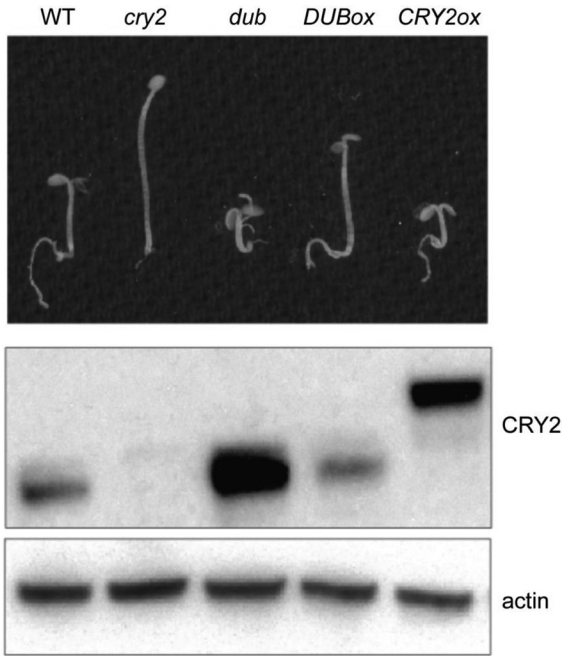


Figure 2. *Arabidopsis* seedling stem (hypocotyl) length is controlled by CRY2 protein levels. Lack or lower levels of CRY2 protein results in a longer hypocotyl as seen in *cry2* mutant and deubiquitinase overexpressor (*DUBox*). Similarly higher CRY2 levels leads to a shorter hypocotyl as seen in the *dub* mutant and CRY2 overexpressor (*CRY2ox*) seedlings.

The Role of Cryptochrome in RNA Metabolism

We also identified many RNA-binding proteins that co-purified with CRY2. Two recent papers also reported co-purification of RNA binding proteins with human CRY1/2. However, the significance of CRYs associating with RNA binding proteins is not known. Interestingly, the CRY2 nuclear speckles resemble those formed by pre-mRNA splicing factors, SR proteins, and other RNA binding proteins in plants and animals, indicating that CRYs likely have a role in RNA metabolism. This observation may provide mechanistic insights into posttranscriptional control, known to be essential for animal and plant circadian biology, and into control of alternative splicing in plants by light. We are currently focusing on two unknown proteins that are hypothesized to bind to

modified RNA with their mutants resembling *cry2* mutant plants, indicating an epistatic relationship between them. Furthermore, similar to our experiments in plants, we are also studying their orthologs in animals to see whether they can also interact directly with mammalian CRY2. RNA modifications are emerging as important regulators of various cellular processes not limited to protein translation, mRNA degradation, alternative splicing, and nuclear export. We are generating loss-of-function mutants of these RNA-binding proteins to understand their impact on plant growth, circadian rhythms, alternative splicing and other physiological responses.

PUBLICATIONS

Conn A, Pedmale UV, Chory J, Navlakha S. 2017. High-resolution laser scanning reveals plant architectures that reflect universal network design principles. *Cell Syst* **5:** 53–62. doi:10.1016/j.cels.2017.06.017

Conn A, Pedmale UV, Chory J, Stevens CF, Navalkha S. 2017. A statistical description of plant shoot architecture. *Curr Biol* **27:** 2078–2088. doi:10.1016/j.cub.2017.06.009

PLANT DEVELOPMENTAL GENETICS

M. Timmermans A. Husbands S. Knauer D. Skopelitis

Boundary Formation through a Direct Threshold-Based Readout of Mobile Small RNA Gradients

D. Skopelitis

Specification of adaxial–abaxial (upper/lower) polarity in the developing leaf drives the flattened outgrowth of the leaf blade and directs the differentiation of distinct cell types within the leaf's adaxial/top and abaxial/bottom domains. Both are important innovations in the evolution of land plants that maximize photosynthesis while minimizing water loss to the environment. In addition to being a key developmental process, adaxial–abaxial patterning is of particular interest from a mechanistic point of view and has proven to be an excellent model to study small RNA–mediated gene regulation. We previously showed that the specification of adaxial–abaxial polarity involves a unique patterning mechanism in which the tasiR-ARF and miR166 small RNAs function as mobile positional signals. Movement of these small RNAs from their defined source of biogenesis in the top and bottom epidermis, respectively, creates opposing small RNA gradients across the leaf, which we showed are interpreted to yield sharply defined expression domains of their targets and polarity in developing primordia (Fig. 1).

To study the mechanism whereby mobility-derived small RNA gradients pattern their targets, we took advantage of artificial microRNA (miRNA) technologies to modulate the position, direction, and steepness of small RNA gradients across the leaf. The outcomes of these experiments show that boundary formation is an inherent property of the small RNA gradient itself. The threshold-based readout of a small RNA gradient is highly sensitive to small RNA levels at the source, allowing plasticity in the positioning of a target gene expression boundary. In addition to generating sharp expression domains of their immediate targets, the readouts of opposing small RNA gradients were found to underlie formation of stable and uniformly positioned developmental boundaries. These patterning properties of small RNAs are reminiscent of those of morphogens in animal systems. However, their unique mode of action and the fully intrinsic nature of their gradients distinguish mobile small RNAs from classical morphogens and present a novel direct mechanism through which to relay positional information. Mobile small RNAs and their targets thus emerge as highly portable and evolutionarily tractable regulatory modules through which to create a pattern in development and beyond.

Dissecting Small RNA Mobility in Plants

D. Skopelitis [in collaboration with K. Hill and S. Klesen, University of Tübingen, Germany]

Given the scope of miRNA-regulated gene networks, the cell-to-cell movement of small RNAs has important implications with respect to their potential as instructive signals in development or response to physiological and stress stimuli. We are using artificial miRNAs targeting easy-to-score reporter genes to study parameters of miRNA movement such as tissue specificity, directionality, dose dependence, and the kinetics of movement. Our findings support the idea that miRNAs traffic via plasmodesmata, small channels that connect most plant cells, and indicate that the number of cells across which a small RNA moves is determined in part by its abundance at the source. Moreover, our findings indicate that the intercellular movement of small RNAs is indeed a developmentally regulated process. Depending on the tissue context, small RNAs can show directional movement or act strictly cell autonomous. Interestingly, production of miR-GFP in young leaf primordia surrounding the shoot apical meristem (SAM) resulted in silencing of GFP inside the SAM, showing that miRNAs traffic from determinate organs into the shoot stem-cell niche. In addition, miR-GFP was able to move from the vasculature below the SAM into the niche.

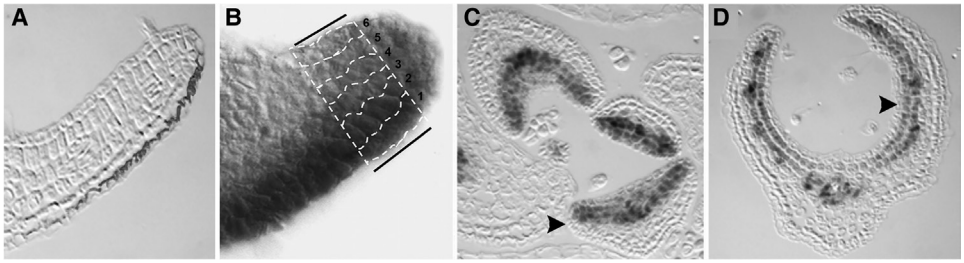


Figure 1. Gradients of mobile small RNAs have morphogen-like patterning activities. Mobility of microRNAs (miRNAs) from their site of biogenesis in the bottom epidermis (*A*) yields an miRNA gradient (*B*) that, through a threshold-based readout, establishes an on–off pattern of target gene expression (*C*,*D*).

These findings suggest that miRNAs may function as mobile signals between differentiated tissues and stem cells, possibly integrating environmental/physiological cues and plant development. A further outcome of this work is that the developmental regulation of small RNA mobility follows rules that are distinct from those that govern protein trafficking. Using the insights gained from the above experiments, we are performing forward genetic screens to identify factors underlying regulated miRNA trafficking. The first screen has yielded several mutants that either increase or decrease the range of miRNA movement from the epidermis, which we are currently analyzing at the molecular level.

The START Domain Regulates HD-ZIPIII Activity and Organ Polarity

A. Husbands

Our previous observations on the expression and function of miR166 and tasiR-ARF reveal that adaxial–abaxial patterning involves a cascade of positional signals. The mobile signals that pattern the newly formed leaf are distinct from those that maintain polarity during subsequent development. This project aims to identify additional signals in adaxial–abaxial patterning. One such signal may involve the adaxial-promoting HD-ZIPIII transcription factors (TFs), which contain a putative lipid-binding START domain. Modeling of the START domain of the HD-ZIPIII member PHABULOSA (PHB) suggests structural similarity to human PC-TP, a phosphatidylcholine-binding START domain. This model, and sequence conservation with other START domains, was used to predict amino acids in PHB critical for START domain ligand binding. Analysis of *Arabidopsis* lines that express PHB-YFP fusion proteins with mutations in key residues lining the START ligand-binding pocket indicates that the START domain is required for proper HD-ZIPIII function. Such mutations block PHB-mediated target gene expression, indicating positive regulation of HD-ZIPIII activity by an unknown ligand.

Experiments to determine the mechanism through which the START domain controls HD-ZIPIII function show that subcellular localization is not impacted by ligand binding. Likewise, ligand binding does not impact the suite of PHB-interacting partners. However, ligand binding does affect PHB's ability to bind DNA, possibly by blocking dimerization. To assess this, we established single-molecule pull-down (SiMPull) in plants, which uses fluorescence-based detection to interrogate protein–protein interactions and can quantitatively determine the heterogeneity and stoichiometry of protein complexes. Using SiMPull, we showed that mutating or deleting the START domain diminishes PHB's capacity to form dimers. Our data further indicate an additional, distinct role for the START domain in augmenting the transcriptional potency of PHB. To independently validate these findings, we fused the START domain to a heterologous TF and found that the behavior of this START fusion protein neatly mirrored that of PHB. Taken together, we propose a model in which the presence of a START domain turns HD-ZIPIII proteins into potent TFs, but only in those cells in which they complex with their ligand. Currently, we are using a number of biochemical approaches to identify the endogenous ligand bound by HD-ZIPIII proteins. The results from immunoprecipitation-mass spectrometry experiments are very encouraging and identify an interesting ligand candidate.

A High-Resolution Gene Expression Atlas for the Maize Shoot Apex

S. Knauer [in collaboration with M. Javelle and M. Scanlon, Cornell University, Ithaca, NY; G. Muehlbauer, University of Minnesota, Minneapolis, MN; J. Yu, Kansas State University, Manhattan, KS; P. Schnable, Iowa State University, Ames, IA]

The SAM, a specialized stem-cell niche at the growing shoot tip, integrates developmental and environmental signals to direct the initiation and patterning of new organs such as leaves. Its activity throughout the plant's lifetime is tightly controlled. To gain insight into gene regulatory networks behind stem-cell maintenance and organogenesis, we generated a high-resolution gene expression atlas of 10 distinct domains and cell types within the vegetative maize shoot apex using laser microdissection and RNA deep sequencing. We found that ~10% of all transcribed genes are differentially expressed across these tissue types, including a valuable collection of cell type–specific genes. Interestingly, very few functional categories are enriched among the differentially expressed genes, which we show reflects prominent subfunctionalization within gene families. Strong enrichments were, however, seen for TF families, and principal component analysis identified unique TF signatures predictive of meristematic and vascular fate. Analysis of TF-binding sites within promoter regions of stem-cell-specific genes predicts a hierarchical network in which the combinatorial actions of diverse TF families underlie their spatially restrictive pattern of expression. Natural variation present at these TFs is associated with key plant architectural traits in genome-wide association studies, providing functional support for these findings. Moreover, through cluster analysis we identified genes whose expression specifically marks the functional zones of the maize SAM—that is, the stem cell–harboring central zone, the organogenic peripheral zone, and the organizing center. Genes defining these zones are in part conserved between maize and *Arabidopsis*, but also reveal remarkable differences and novel gene functions associated with these domains in maize that are validated by mutational analysis. In summary, our findings identify unique TF signatures as master regulators of cell identity within the SAM that balance stem-cell maintenance and organogenesis.

PUBLICATIONS

Skopelitis DS, Benkovics AH, Husbands AY, Timmermans MCP. 2017. Boundary formation through a direct threshold-based readout of mobile small RNA gradients. *Dev Cell* **43:** 265–273.

GENOMICS

There has been a growing appreciation in recent years that gene function is frequently context dependent, with a large part of that context provided by the activities of other genes. But trying to understand how genes interact to produce function is a hugely complicated problem and one that appears likely to become more so as genomic information becomes more detailed. **Jesse Gillis** and colleagues are computational biologists who are presently challenging an oft-taken approach to the problem in which the functions of genes are interpreted in the context of networks derived from gene association data. Such networks consist of millions of interactions across thousands of genes, derived from protein-binding assays, RNA coexpression analysis, and other sources. Historically, many attempts to understand gene function through networks have leveraged a biological principle known as "guilt by association." It suggests that genes with related functions tend to share properties (e.g., physical interactions). In the past decade, this approach has been scaled up for application to large gene networks, becoming a favored way to grapple with the complex interdependencies of gene functions in the face of floods of genomics and proteomics data. Gillis' work centers on identifying the limits of the approach and making fundamental improvements to its operation, as well as applying those improvements to neuropsychiatric gene network data.

Thomas Gingeras and colleagues study where and how functional information is stored in genomes. These efforts help explain the biological and clinical effects of disease-causing gene mutations in humans and other organisms. Gingeras is a leader of the ENCODE (ENCyclopedia of DNA Elements) and the mouseENCODE and modENCODE (model genome ENCODE) projects of the National Institutes of Health. His research has altered our understanding of the traditional boundaries of genes, revealing that almost the entire lengths of genomes in organisms ranging from bacteria to humans can be transcribed into RNA (pervasive transcription), and most RNA products made by a cell are not destined to be translated into proteins (noncoding RNAs, or ncRNAs). In fact, ncRNAs are proving to be involved in a variety of other important biological functions. Some have been shown to be critical components in the pre- and posttranscriptional and translational processes, as scaffolds on which large protein complexes are assembled and as extracellular signals. The initial studies that led to these observations have been extended to cover the entire human genome.

Gholson Lyon's lab focuses on analyzing human genetic variation and its role in severe neuropsychiatric disorders and rare diseases, including Tourette syndrome, attention-deficit hyperactivity disorder (ADHD), obsessive compulsive disorder (OCD), intellectual disability, autism, and schizophrenia. By recruiting large groups of related individuals living in the same geographic location (e.g., Utah), Lyon's lab can study the breadth and depth of genetic variants in a similar environmental background. Using the exome—the parts of the genome that code for protein—and whole-genome sequencing, his lab looks for mutations that segregate with syndromes in the various populations. A second focus of the Lyon lab is to study the mechanistic basis of a new rare disease that they described in 2011. This is the first human disease involving a defect in the amino-terminal acetylation of proteins, a common modification of eukaryotic proteins performed by amino-terminal acetyltransferases (NATs). The team has been using several different cellular model systems to better understand the disease pathophysiology and the basic process of amino-terminal acetylation. This year, Lyon collaborated with a team of researchers from other universities and companies to use precision medicine to successfully treat a patient with severe OCD. His symptoms were treated with deep brain stimulation, and the team used whole-genome sequencing to try to understand the molecular basis of his disease. The patient experienced significant relief

from his symptoms and his quality of life returned, suggesting that similar methods may hold tremendous promise in the future.

The insights of **W. Richard McCombie** and colleagues have led to the introduction and optimization of novel methods of high-throughput genome sequencing. His team has made it possible to catalog variation among individual organisms in a way that would have been unthinkable 10 years ago. They have brought online a new generation of Illumina sequencers and optimized their function to a level at which eight to 10 trillion DNA bases can be sequenced in a month. McCombie's team has been involved in international efforts culminating in genome sequences for maize, rice, and bread wheat—three of the world's most important food crops. They have also had an important role in projects to sequence the flowering plant *Arabidopsis thaliana* (the first plant genome sequence), the fission yeast *Schizosaccharomyces pombe*, the human genome, and other important genomes. McCombie's group is currently involved in several important projects to resequence genes in patient samples that are of special interest to human health, including *DISC1* (a strong candidate gene for schizophrenia), looking for genetic variants implicated in bipolar illness and major recurrent depression. They are also looking for genes that contribute to cancer progression using whole-genome sequencing or a method called exome sequencing, which they developed with Greg Hannon to look at mutations in the regions of the genome that code for proteins.

Using multidisciplinary approaches that combine computational analysis, modeling, and prediction with experimental verification, **Doreen Ware**'s lab seeks a deeper understanding of the evolution of genome sequences in plants and their implications for agricultural improvement. By looking comparatively across the genomes of plants in the same lineage, they seek answers to the following questions: How are genes conserved and lost over time? What are the fates of duplicated genes? What is the impact of structural variation on phenotypic variation? Ware's team also studies gene regulation in plants, focusing on gene regulatory networks, targeting transcription factors and microRNA genes with the objective of understanding how these parts of the plant genome work together in determining spatial and temporal expression of genes. The lab had an important role in the project to produce a haplotype map reference genome of maize, spearheading the most comprehensive analysis of the crop yet. This has provided important information on the variation of the reference genome, as well as comparative data showing changes in the genome acquired through domestication and breeding. They have devoted special attention to examining diversity within maize, grape, and tomato, aiming to accelerate the development of strategies to introduce new germplasm that is needed to meet demands of increasing population and a changing environment. The lab also has brought fully sequenced genomes into an integrated data framework to enhance the power of their comparative studies. This past year, Ware was named principal investigator for the National Science Foundation–funded Gramene project, a comparative genomics resource for agriculturally important crops and models to support sustainable food and fuel production. Ware, as principal investigator for plants, has also helped lead an effort funded by the Department of Energy to create—out of many separate streams of biological information—a single, integrated cyber-"knowledgebase" for plants and microbial life.

GENE EXPRESSION, FUNCTION, AND META-ANALYSIS

J. Gillis S. Ballouz M. Crow M. Shah

Research in the Gillis lab involves two interwoven elements: improving the interpretability of network analysis and characterizing transcriptional data in the brain. These topics form a naturally complementary unit because the complexity of the brain as a system means that it is essential that the methods for analyzing it yield clear and precise signals. A dominant interest within computational biology is the analysis of gene networks to provide insight into diverse levels of functional activity, typically starting with regulatory interactions and moving up to more diffuse associations important for understanding systemic dynamics. Gene associations (of various sorts) are believed to encode functional interaction, and this interaction is frequently shown to be able to substantially predict gene function. This approach, commonly called "guilt by association," is embedded in everything from prioritization of de novo variants to uncovering novel regulatory interactions or mechanisms of disease. Although black box style network analyses are common, explaining the basis of how and why methods work is more rarely attempted. In the Gillis lab, we are developing network-based methods and software that improve both the sophistication and breadth of data available for determining how genes interact to produce function, particularly focusing on how genes interact to cause disease or cell phenotypes. Broadly, our research can be divided into methods development and our own research applications, often performed in collaboration with other labs to test computational predictions experimentally. In addition to Jesse Gillis, the members of the Gillis lab are Postdoctoral Fellows Sara Ballouz and Maggie Crow and Computational Science Developer Manthan Shah.

Understanding Drivers of Differential Expression

Differential expression (DE) is commonly used to explore molecular mechanisms of biological conditions. Although many studies report significant results between their groups of interest, the degree to which results are specific to the question at hand is not generally assessed, potentially leading to inaccurate interpretation caused by base rate neglect. To address this, we aimed to determine whether any gene signatures are commonly differentially expressed within a comprehensive database of human microarray and RNA-sequencing (RNA-seq) data. Our approach is to generate an "optimal prior" that uses the likelihood of a gene to be DE to predict the results of individual studies in the absence of any specific knowledge about the underlying biology. We find that the optimal prior predicts DE genes in nearly all experiments with high accuracy, supporting the existence of base rate neglect within genomics. We then identify 175 genes that are of high probability to be differentially expressed across multiple expression platforms, and find that they are linked to three fundamental biological processes: sex, inflammation, and development. Using sex as a model case, we show that controlling for these factors within statistical analyses has a large impact on the significance of results and on their interpretation. We suggest that future experiments should either be designed to account for these effects or assess their impact and provide easy-to-use tools to aid in both tasks.

RNA-seq Alignment

A major computational challenge in RNA-seq is the quantification of gene expression. At present, the most commonly used approach consists of mapping RNA-seq reads to a reference, followed by calculation of the transcript or gene expression. Many tools are available for RNA-seq alignment and expression quantification—with comparative value being hard to establish. Benchmarking assessments often highlight methods' good performance with occasional outlier tools but either are focused on model data or fail to detail variation in performance. This leaves us to ask, what is the most meaningful way to assess different alignment choices? And importantly, where is there room for progress? In this work, we explore the answers to these two questions by performing an exhaustive assessment of the STAR aligner in collaboration with Alex Dobin

and Tom Gingeras. We assess STAR's performance across a range of alignment parameters using common metrics, and then on biologically focused tasks. We find technical metrics such as fraction mapping or expression profile correlation are uninformative, capturing properties unlikely to have any role in biological discovery. Surprisingly, we find that changes in alignment parameters within a wide range have little impact on both technical and biological performance. Yet, when performance finally does break, it happens in difficult regions, such as X Y paralogs and major histocompatibility complex (MHC) genes. We believe improved reporting by developers will help establish where results are likely to be robust or fragile, providing a better baseline to establish where methodological progress can still occur.

Expression Outliers in Rare Disease

Assessing gene expression from individual patients in rare disease cohorts is a statistical problem—we have little power to detect changes in a robust fashion. One typical way to overcome this $n = 1$ problem is to gain more power by looking for a joint signal among differentially expressed genes through their shared functions. These functional signals are usually molecular pathways to which the genes belong or gene ontology (GO) terms they share. An alternative to this approach is to look at the genes that should be frequently coregulated, but now display an unusual expression signature and, thus, potentially harbor a disease signal that is both disruptive and unique. In this project, we examine the possibility that the candidate disease genes display "outlier" or unexpected expression rather than reflecting a collective pattern of dysregulation.

In collaboration with the Lyon lab, we collected and sequenced one quad and five trios from the rare *TAF1* syndrome cohort. The probands from each family have a unique variant in their *TAF1* transcription factor, and share many phenotypic attributes. As we are looking for unusual DE in a disease context, we were able to design a novel family-specific DE analysis that exploits expected overlaps and differences in the transcriptomic profiles of the parents relative to their affected child. Then, by tallying coexpression patterns from more than 3000 expression samples in 75 experiments, we generated a frequency of common expression value for all gene pairs across most of the genome. Genes with jointly common expression values were then filtered away, leaving us with a small number of what we called outlier genes, characteristic of each proband. We first note that filtering removed all GO functional enrichment. Importantly, we found a single outlier gene, the calcium channel subunit *CACNA1I*, that recurs in five of the six pedigrees. Notably, this gene is recurrently implicated in other neurological diseases, such as schizophrenia and autism, making it a very plausible candidate. The sole family in which no signal was present was a copy number variant (CNV) carrier, whereas the other probands had different single-nucleotide variants, implying a potentially different underlying molecular mechanism. Our outlier-based analysis revealed otherwise difficult-to-observe signals in the case of *TAF1* syndrome cohort, and we anticipate that future transcriptomic studies of rare disorders would benefit from this type of analysis.

Learning Cell Identity

The exceptional diversity of neurons has been appreciated from the time of Ramón y Cajal. His depictions of cells that varied wildly in size, shape, and connectivity provided foundational guidance in the field of neuroscience. The precise extent of this diversity has been debated since that time, with each new technology indicating novel facets of neuronal identity: from morphology, to electrical activity, and now single-cell transcriptomics. When faced with a transformative technology like single-cell RNA-seq, how are we to make sense of the varied and distinct results of papers attempting to reshape the landscape of molecular neuroscience? One approach is to build on a firm bedrock of our prior knowledge; comparison across studies to detect replicability is another. Our ongoing work with the Huang lab combines these two ideas in a naturally complementary way: By assessing clusters of cells for their overlap across studies with respect to the known functional properties of the genes that describe them, we characterize not just the overlap in transcriptional signal, but also its functional implications. These innovations, and our method of cross-laboratory study design, are critical means of defining the overall state of neuroscientific knowledge as to cell identity.

We first measured the replicability of neuronal identity by comparing more than 13,000 individual

scRNA-seq transcriptomes, and then assess cross-dataset evidence for novel pyramidal neuron and cortical interneuron subtypes identified by scRNA-seq. We found that 24/45 cortical interneuron subtypes and 10/48 pyramidal neuron subtypes have evidence of replication in at least one other study, and provide lists of candidate marker genes. Across tasks we found that any large set of variably expressed genes can identify equivalent cell types across datasets with high accuracy, indicating that many of the transcriptional changes defining cell identity are pervasive and easily detected.

Functional Convergence in Autism

Disagreements over genetic signatures associated with disease have been particularly prominent in the field of psychiatric genetics, creating a sharp divide between disease burdens attributed to common and rare variation, with study designs independently targeting each. Meta-analysis, including using underlying data, within each of these study designs is routine, but no method for combining the results across study designs exists. Traditional meta-analytic approaches are not calibrated to detect functional convergence within this data because significant disease–gene associations from the diverse study designs arise in quite different ways, with quite different null hypotheses. In this work, we develop a general solution that integrates the disparate genetic contributions, constrained by their observed effect sizes, to determine functional convergence in the underlying architecture of complex diseases, which we illustrate on autism spectrum disorder (ASD) data.

Our approach looks not only for similarities in the functional conclusions drawn from each study type individually but also those that are consistent with the known effect sizes across these studies. We name this the "functional effect size trend" and it can be understood as a generalization of a classic meta-analytic method, the funnel plot test. Because this allows us to combine many different studies, and do so in a structured way, it has the potential to find overlapping functional signals that would otherwise be impossible to see. We took candidate disease gene data from multiple ASD studies across thousands of individuals (~7000) and study designs, including whole-exome sequencing and genome-wide association studies. We split the candidate genes by variant class (common and rare) and effect size (low to high) into 14 gene sets in the range of 20–100 associated genes per set, and performed functional analyses, controlling for set size. We detected remarkably significant trends in aggregate, with 20 individually significant properties (false discovery rate [FDR] <0.01)—many in areas researchers have targeted based on different reasoning, such as FMRP interactor enrichment (FDR of ~0.006). We are also able to detect novel technical effects and we see that network enrichment from protein–protein interaction data is heavily confounded with study design, arising readily in control data. We perform a number of downstream analyses, including tests for robustness and genome-wide rankings of disease association. Our focus is on ASD and gene-level associations, but the method can be extended to other diseases and other types of tests as data permits (e.g., phenotype screening, grouping of disorders, variants within studies).

Epigenetic Meta-Analysis

Chromatin accessibility provides an important window into the regulation of gene expression. Recently, the assay of transposase accessible chromatin with sequencing (ATAC-seq) was developed to profile genome-wide chromatin accessibility. Although pipelines for implementation have been developed on an ad hoc basis for the analysis of individual data sets, there has been little comparative or aggregate evaluation. This is critical both to determine appropriate methodologies, controls, and efficacy, as well as to determine the global biological landscape of chromatin accessibility across diverse conditions. One major technical problem to address is that the counts of ATAC-seq reads underlying each peak vary substantially within a single sample and also between samples. Such variation makes the comparison to determine presence and absence of peaks (i.e., the open and closed state of chromatin) more difficult and less statistically well grounded. In this work, we analyzed 197 ATAC-seq data samples from 13 studies to evaluate the robustness of results, as well as their specificity across studies. We find that peaks are promiscuously identified, with approximately 34,000 peaks per sample on average. These substantially overlap with transcription start sites (TSSs), with 11,000 genes on average overlapping with the called peaks. Among those genes, 447 genes

have peaks at their TSSs from all 197 samples we analyzed. We evaluate the properties of these genes in detail, including mean expression across a diverse corpus of data, coexpression between this set and other genes, and functional enrichment. Finally, we suggest a novel approach to evaluate the robustness of peak signals and sensitivity by bootstrapping reads and re-calling peaks for each resampling, whose calls are then aggregated. This yields peak calls that are highly robust to variation in noise as a source of peaks within data itself. We applied this novel data to our own ATAC-seq conditional experiment, with a reduction of likely spurious peaks to improve specificity.

Heuristics for the Interpretation of Gene Sets

One branch of our research tackles an almost philosophical problem of broad impact in bioinformatics: How can we evaluate methods in which no gold standard exists? This is a particular problem when the bioinformatics methods are targeted to functional interpretation, in which not only is a gold standard challenging, but knowing whether method performance will generalize is a major impediment. The most striking example of this problem occurs in function enrichment methods, which are ubiquitous, diverse, and very hard to assess.

We propose a solution to this problem by looking to assess not whether enrichment results are correct, but whether they meet basic criteria for coherence. In particular, we show that assessing algorithm outputs for uniqueness and robustness is highly useful across a variety of problems. A substantial part of the work involves formalizing a test for uniqueness or robustness that can be applied to any enrichment method, but ultimately the principles map to surprisingly intuitive biological properties in the form of an assessment of the role of multifunctional genes within study results.

Gene set analysis to translate gene lists into enriched functions is among the most common bioinformatic methods, yet few would advocate taking observed p values at face value. Not only is there no agreement on the algorithms themselves, there is no agreement on how to benchmark them. In this paper, we evaluate the robustness and uniqueness of enrichment results. We show that both properties turn out to be closely linked to multifunctionality, defined as the number of annotated functions a gene has. We show that multifunctional genes are more likely to appear in genomics study results and drive the generation of biologically nonspecific enrichment results because of their overlaps in annotations. Additionally, individual multifunctional genes easily generate apparent, yet highly fragile, significance. By providing a means of benchmarking when enrichment analyses report nonspecific and nonrobust findings, we are able to assess where we can be confident in their use. We find that gene set enrichment methods that correct for biases are far likelier to return robust and unique results. We provide a software implementation for multifunctionality assessment in ErmineJ, but our general approach can be adapted to any enrichment method.

PUBLICATIONS

Ballouz S, Gillis J. 2017. Strength of functional signature correlates with effect size in autism. *Genome Med* **9:** 64.

Ballouz S, Pavlidis P, Gillis J. 2017. Using predictive specificity to determine when gene set analysis is biologically meaningful. *Nucleic Acids Res* **45:** e20.

Ballouz S, Weber M, Pavlidis P, Gillis J. 2017. EGAD: Ultra-fast functional analysis of gene networks. *Bioinformatics* **33:** 612–614.

Paul A, Crow M, Raudales R, He M, Gillis J, Huang ZJ. 2017. Transcriptional architecture of synaptic communication delineates GABAergic neuron identity. *Cell* **171:** 522–539.e20.

In Press

Crow M, Paul A, Ballouz S, Huang ZJ, Gillis J. 2018. Characterizing the replicability of cell types defined by single cell RNA-sequencing data using MetaNeighbor. *Nat Commun* **9:** 884.

Kalish BT, Cheadle L, Hrvatin S, Nagy MA, Rivera S, Crow M, Gillis J, Kirchner R, Greenberg ME. 2018. Single-cell transcriptomics of the developing lateral geniculate nucleus reveals insights into circuit assembly and refinement. *Proc Natl Acad Sci* **115:** E1051–E1060.

GENOME ORGANIZATION, REGULATION, AND FUNCTIONAL ROLES OF NONCODING RNAs

T.R. Gingeras C. Danyko G. Nechooshtan
A. Dobin S. Patel
J. Drenkow D. Yunusov

Human-ENCODE Project Transition to ENT-Ex Project

C. Danyko, A. Dobin, J. Drenkow

Phase 3 of ENCODE (3) continues in the form of the ongoing collaboration the National Human Genome Research Institute (NHGRI)-funded ENCYCLOPEDIA of DNA Elements (ENCODE) and the National Institutes of Health (NIH) Common Fund–funded Genotype-Tissue Expression (GT-Ex) projects. All of the data generated for the ENT-Ex and the transcriptome portions of the ENCODE3 project are available at https://www.encodeproject.org/search/?type=experiment&lab.title=Thomas%20Gingeras,%20CSHL. As its core goal, the EN-TEx project involves deriving a set of precision genomes for four individuals (two male, two female). To accomplish this there has been an effort to determine not only the genome sequences of the four donors but also RNA-Seq data from 20 tissues per donor and epigenetic profiles from each of the tissues collected. These data have been collected to construct highly detailed personal phased diploid genomes for each donor and correlate the genetic variations found in each genome with the tissue transcriptional and epigenetic profiles. To achieve this integration of data types, multiple sequence and sequence analysis software were used for each genome, transcriptome, and epigenome. Short-read Illumina genome sequence data for each of the four donors was collected at a depth of >100× coverage. In addition, phasing of the genomes as determined using HiC sequence analysis and phased diploid genomes were constructed by use of long-read PacBio and 10× analyses. Unique and common single-nucleotide polymorphisms (SNPs) and structural variations for each genome were determined for each allele of each genome. Gene expression analyses for each tissue from each donor were performed for long (>200 nt) and short (<200 nt) RNAs, and gene expression was quantified in multiple tissues using RAMPAGE (promoter-specific RNA sequencing). ChIP-seq analyses were performed for several activation (H3K4me1, H3K4me3, H3K27ac, H3K36me3) and repression (H3K9me3, H3K27me3) chromatin marks, as well as Pol2, EP300, and CTCF transcription factors. DNase hypersensitivity analysis for each of the four genomes was performed to assist in constructing unsupervised segmentation of the genomes into multiple chromatin states, creating spatially resolved combinatorial regulatory maps in a tissue-specific manner. The states are annotated as various types (active, inactive, bivalent) of promoters, enhancers, and transcripts, using prior knowledge as well as overlap with known genomic features. These constructed segmentations were then correlated with the transcription and ChIP-seq signals to further refine the regulatory maps. Unlike the standard RNA-Seq assay, which quantifies the expression of entire transcripts/genes, RAMPAGE detects and quantifies transcription start sites, thus allowing for direct comparison between ChIP-seq and RNA signals on promoters. The active promoter regions defined in the unsupervised segmentation of ChIP-seq data are strongly enriched for RAMPAGE (promoter) peaks, attesting to the high fidelity of chromatin segmentation. We built a neural network model for predicting the promoter expression from the ChIP-seq signals and found that ~65% of the promoter expression variation can be explained by the ChIP-signal variation, with the main predictors being H3K4me3, H3K4me1, and Pol2. The tissue-specific functional genome segmentations are used for annotating the variants contributing to variation in gene expression (expression quantitative trait loci [eQTLs] identified by the GT-Ex consortium). This resulted in a significant statistical enrichment for genome-wide association study (GWAS) and eQTL variants in chromatin

states interpreting the disease-/trait-associated variants detected in GWAS studies. Finally, the effectiveness of the segmentation results was evaluated by determining whether tissue specificity for hypercholesterol phenotypes could be observed. Using data from 83 independent genotype–phenotype studies, a total of 1183 GWAS SNPs were reported in 16 tissues. Liver tissue was observed to have the most significant number of SNPs, with SNPs overlapping segmentation called active promoters. Results like these underscore the value of replacing the currently used reference human genome sequence (v38) with individual personal genome sequences.

Database Management and Software Development

A. Dobin

Personal genomics is envisaged to become an essential component of precision medicine, holding the promise for identifying genetic predispositions for common and complex diseases, diagnosis of hereditary disorders, individual treatment of cancer, and genotype-guided drug research. Millions of personal genomes will be sequenced in the next few years; however, some software tools are lacking for personalized processing of the functional data types such as RNA-Seq and ChIP-seq, which at present are routinely mapped to the haploid reference genome. To address this need, an extension of popular RNA-Seq aligner STAR, the STAR-Diploid software, was developed to map RNA- and ChIP-seq reads to the fully phased diploid personal genomes.

The development of STAR-Diploid software has been completed. This software uses the personal variants, including SNPs, short indels, and large structural variants, to build the personal diploid genome sequence from the reference assembly. The reference annotations are arithmetically lifted over to each of the haplotypes. Next, the reads are mapped to both haplotypes simultaneously to produce diploid genomic alignments. Mapping to the personal diploid sequence virtually eliminates the reference bias that plagues the alignment to the haploid reference genome. The diploid alignments are then converted to the reference coordinates while preserving the haplotype information. The final output of the pipeline consists of haplotype-specific alignments and signal (wiggle) tracks in the reference coordinates, which can be visualized in the standard genomic browsers, as well as allele-specific counting of reads per gene. Furthermore, STAR-Diploid converts diploid genomic alignments into diploid transcriptomic alignments, which are input into RSEM for allele-specific quantification of transcripts and gene expression.

To show the effectiveness of the STAR-Diploid algorithm, we have used it to process a large collection of long and small RNA-Seq, RAMPAGE, and ChIP-seq data from the EN-TEx project (approximately 20 tissues for four donors). The personal diploid genomes were constructed with the variants phased by means of 10× Genomics Chromium sequencing and Hi C short reads for chromosome-span phasing, with RNA-Seq data used to supplement and resolve phasing conflicts. We have compared allelic imbalance across the multiple tissues obtained from the same donor, as well as across multiple individuals for the same tissue, for the purpose of understanding the genotypic and cell type contributions to allele-specific expression (ASE). Using long-range variant phasing information from the 10× Genomics and Hi-C data, we have identified candidate causative mutations in regulatory regions responsible for the observed ASE.

Sorting and Processing of Various Types of RNAs in the Extracellular Environment

G. Nechooshtan, S. Patel

RNA is not only found within the boundaries of cells. We had turned our attention to the identification of a collection of RNAs that undergo site-specific processing in the extracellular milieu. Examples of these RNAs include tRNAs and Y RNAs. One of our first goals of this project is to identify and isolate the factors involved in this processing activity. Following proteomic and biochemical work, we have identified RNase 1 as a possible candidate for this activity.

During 2017, we have shown that addition of commercial RNase inhibitor to tissue culture medium inhibits the processing of RNY5 found in the medium of cells grown in culture. RNase inhibitor specifically inhibits RNase 1 family enzymes. Because RNase 1 is the most highly expressed and most catalytically active member of the family, a modified cell line was generated that carries an inactivated *Rnase1* gene using a CRISPR-Cas9 approach.

Following the generation and isolation of this mutant cell line, we showed that this mutant shows a greatly reduced processing activity toward secreted RNY5. This supports the candidacy of RNase 1 as a critical factor in RNY5 processing as well as possibly in processing of a wider selection of RNA biotypes. We are currently assessing the scope of this phenomenon.

Because RNase 1 is a secreted enzyme and endogenous RNase inhibitor is abundant inside cells, it is likely that RNase activity is limited to the extracellular environment. However, there may be intracellular compartments or niches in which RNase 1 exists and is protected from inhibition by endogenous RNase inhibitor. We therefore intend to investigate the effect of inactivation of RNase 1 on cellular transcripts as well as secreted transcripts within or outside extracellular vesicles.

Biological Importance of EVs and Their Functional Importance

Updates on phenotypes associated with K562 EVs

D. Yunusov

In the past year, we have attempted to validate and reproduce previously observed phenotypes reported for extracellular vesicles (EVs) and the RNY5 cargo. Specifically, we have previously observed that EVs from K562 human chronic myelogenous leukemia cells (K562 EVs), as well as total RNA isolated from these EVs (K562 EV total RNA), were capable of increasing cell death in human normal foreskin fibroblasts (BJ cell line). We have been able to attribute this phenotype to the most abundant RNA found in K562 EVs—RNY5, which in EVs is predominantly present as a 29–31-nt-long 5′-fragment of the full-length RNY5. By using a variety of other cell lines, we have also observed increased cell death in a subpopulation of target cells as a general response of nontumor human cell lines to EVs from human tumor cell lines. In our current experiments, the cell death phenotype of BJ cells in response to RNY5 31-mer, K562 EV total RNA, or K562 EVs was determined to be part of a general cellular stress response shown by a subpopulation of the treated cells and not an end-point phenotype for all treated cells. More recently, analysis of mouse EVs noted a death response of normal mouse embryonic fibroblasts (NIH/3T3 cell line) to EVs from two mouse breast cancer cell lines (EMT6 and Eph4 1424.1). It should be noted, however, that mouse genome does not encode a *RNY5* gene.

Reanalysis of the RNA-Seq data for human umbilical vein endothelial cells (HUVECs), which were treated with K562 EVs or transfected with RNY5 31-mer, showed high variability across replicates and did not allow for consistent identification of differentially expressed genes. Similar reanalysis of global transcriptional response of BJ cells to K562 EVs, K562 EV total RNA, or RNY5 31-mer did not indicate previously observed enrichment of differentially expressed genes in transforming growth factor β (TGF-β) pathway, but, instead, we could see differential expression of genes involved in cell migration as a response of BJ cells to both K562 EVs and K562 EV RNA but not to RNY5 31-mer. Differential expression of 12 out of 12 randomly selected genes in BJ cells in response to K562 EVs was validated with reverse transcription quantitative polymerase chain reaction (RT-qPCR), and migratory behavior of these cells was confirmed with scratch-wound assay.

It is possible that such disagreement of the more recent results with previous observations, at least in part, can be explained by high variability in EV and EV RNA yields from different biological replicates. We, therefore, have attempted to implement more controlled EV and EV RNA preparation methods. Some of the modifications to the previously described EV isolation method include an every-day control of cell numbers used for medium conditioning for EV isolation, normalization of EV input for downstream experiments by the EV total protein content, and comparisons with ultracentrifugation-based methods. Comparison between the filtration-with-subsequent-precipitation method and the ultracentrifugation method was performed by sequencing total RNA content of K562 EVs obtained with each method. Results showed good correlation between samples and the differences are currently being examined in detail. We have also optimized the EV RNA isolation protocol to decrease the number of steps that can result in loss of EV RNA material.

Finally, although our efforts using cell culture systems have been informative, a more biologically relevant system would be required to study cell-to-cell communication via EVs. One approach to address this challenge has been initiated and involves collecting EVs and total EV RNA from mouse embryonic stem cells and trophoblast stem cells. This would allow us

to model and understand how the cells of the two embryonic lineages—inner cell mass and trophoblast—communicate with each other to establish proper developmental timeline and patterns.

Development of new phenotypic assays for EVs and EV RNAs as a means of detection of clinical cancer

S. Patel

Cell migration is an important step in the metastatic process. Cancer cells degrade their nearby extracellular matrix components and create migration channels. Several studies suggest that EVs secreted by cancer cells contain different repertoires of proteins, small RNAs, and lipids compared with normal cells' EVs. To study the effect of cancer cell EVs on cell motility, EVs were collected from normal and breast cancer cell lines. To study the influence of EVs on cell migration, a trans-well assay system was used to measure the number and proportion of migratory cells treated with cancer cell EVs. The trans-well membrane has two compartment systems, consisting of upper insert and lower chambers, that are separated by an 8-µm membrane. MCF10A nontumorigenic epithelial cell lines were used as target cells for the EVs. In preparation for in vivo studies involving patient plasma EV samples obtained from cancer patients, several variables of the trans-well assay

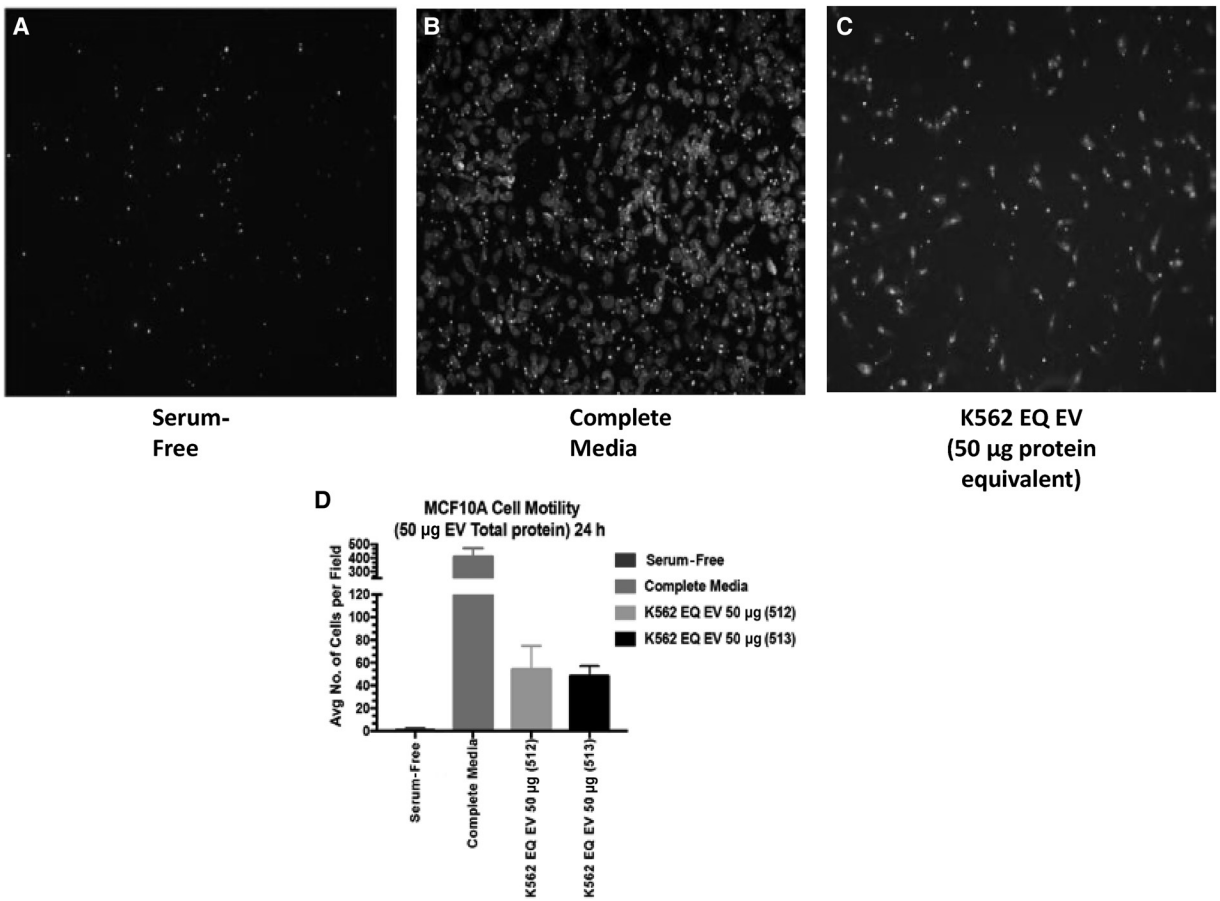


Figure 1. Quantitative cell migration assay to measure the phenotypic effect of EVs. (*A*) MCF10A cells that migrate through 8-µm pores to the underside of the trans-well filter as a response to serum-free cell culture medium, which is the medium used for experiments observed in *C*. (*B*) Same cells that migrate when grown in serum containing medium supplemented with growth factors, (*C*) MCF10A cells grown in serum-free medium supplemented with 50 µg equivalents of K562 EVs. (*D*) Average number of cells that migrate. These values represent the average number from each of five representative microscopic fields per replica (two replicas/sample, S12 and S13) and are seen in each of the experiments.

were evaluated. These variables include methods of EV isolation, minimum number of EVs required to see the effect, and reproducibility of assay using EVs collected from normal and cancer cell lines. Briefly, the MCF10A cells are plated on the upper surface of the trans-well inserts. After 24-h or 48-h periods, activated cells have been observed to pass through the 8-μm pores and attach to the other side of the membrane. The membrane is then stained using fluorescent live cell dye, and removed and then mounted on a glass slide. For cell counting, five representative areas from each membrane surface are inspected and the cells counted. The EVs from cells grown in serum-free medium (negative control) (Fig. 1A) and cells grown in media containing growth factors (positive control) (Fig. 1B) are counted and compared with cells treated with K562 EVs (Fig. 1C,D). In summary, ~3%–5% of the EV-treated (50 μg of EV-protein) MCF10A cells migrate, which is 15–25-fold greater than the negative controls. This assay will provide a quantitative complement to the in vivo experiments that are ongoing.

PUBLICATIONS

Batut PJ, Gingeras TR. 2017. Conserved noncoding transcription and core promoter regulatory code in early *Drosophila* development. *Elife* pii: e29005.

Breschi A, Gingeras TR, Guigó R. 2017. Comparative transcriptomics in human and mouse. *Nat Rev Genet* **18:** 425–440.

Lagarde J, Uszczynska-Ratajczak B, Carbonell S, Pérez-Lluch S, Abad A, Davis C, Gingeras TR, Frankish A, Harrow J, Guigo R, et al. 2017. High-throughput annotation of full-length long noncoding RNAs with capture long-read sequencing. *Nat Genet* **49:** 1731–1740.

Noguchi S, Arakawa T, Fukuda S, Furuno M, Hasegawa A, Hori F, Ishikawa-Kato S, Kaida K, Kaiho A, Kanamori-Katayama M, et al. 2017. FANTOM5 CAGE profiles of human and mouse samples. *Sci Data* **4**: 170112.

HUMAN GENETICS AND GENOMIC MEDICINE

G. Lyon J. Crain A. Ismail T. Papazyan A. Solowinska
M. Doerfel R. Kleyner A. Sebold Y. Wu

The Amino-Terminal Acetylation of Proteins

M. Doerfel, Y. Wu, J. Crain [in collaboration with J. Lupski and L. Meng, Baylor University; J. Wu, Stanford University; T. Arnesen, Norway; G.T. Oh, Korea; J. Gecz, Australia]

We previously identified the genetic basis of an X-linked, infantile lethal Mendelian disorder, involving a p.S37P missense mutation in *NAA10*, encoding the catalytic subunit of NatA, involved in amino-terminal acetylation (NTA) of proteins. To date, there are more than 20 individuals with damaging mutations in *NAA10* or *NAA15,* with the latter gene encoding the dimeric binding partner for Naa10. Both *NAA10* and *NAA15* are relatively intolerant of variation. Phenotypic features of patients carrying disruptive variation in NatA include craniofacial anomalies, hypotonia, global developmental delays, cardiac arrhythmia, and/or cardiomyopathy. Functional analysis of the *NAA10* missense mutations has shown impaired biochemical activity and/or a reduced capacity to form a stable NatA complex, although there is no proteome-wide decrease in NTA, thus implying tissue-specific and/or temporal substrate-specific effects. Attempts to knock down or knock out *NAA10* in human cells result in cell death, making it difficult to assess proteome-wide NTA in human KO cells. Mice with decreased or absent *Naa10* survive throughout embryogenesis but display a range of phenotypes that are variably expressive and include increased early neonatal lethality (likely related to congenital heart defects), supernumerary thoracic ribs and vertebrae, piebaldism, and/or urogenital and renal abnormalities. Some of these phenotypes overlap with those seen with the missense mutation in humans, but survival of the KO mice was unexpected, given the mutation intolerance of this gene in humans. Extensive proteomic analysis of mouse embryonic fibroblasts has not shown any proteome-wide decrease in NTA in the KO cells, with one possible explanation being our discovery of a previously unannotated *NAA10*-like gene in mice, which lacks a human ortholog, and which we refer to as *NAA12*. A rabbit polyclonal antibody raised with specificity to Naa12 shows expression of Naa12 in many mouse tissues. We used CRISPR to create indels in this new putative *NAA12*, and we are currently characterizing these KO mice, along with breeding them to make the double KO for *NAA10* and *NAA12*. Given the potential redundancy in mice, fibroblasts from one affected human proband control were reprogrammed into induced pluripotent stem cells and differentiated into cardiomyocytes (CMs). Electrophysiological experiments show rate-dependent prolongation of the effective refractory period in patient CMs, suggesting a cell-autonomous defect in the CMs.

The auxiliary subunit of the NatA complex, NAA15, is the dimeric binding partner for NAA10. Through a genotype-first approach with whole-exome or -genome sequencing (WES/WGS) and targeted sequencing analysis, we identified and phenotypically characterized 37 individuals from 32 unrelated families with 25 different de novo or inherited, dominantly acting likely gene disrupting (LGD) variants in *NAA15*. Clinical features of affected individuals with LGD variants in *NAA15* include variable levels of intellectual disability, delayed speech and motor milestones, and autism spectrum disorder. Additionally, mild craniofacial dysmorphology, congenital cardiac anomalies, and seizures are present in some subjects. RNA analysis in cell lines from two of the patients showed degradation of the transcripts with LGD variants, likely because of nonsense-mediated decay. Functional assays in yeast confirmed a deleterious effect for two of the LGD variants in *NAA15*. We propose that defects in NatA-mediated NTA lead to variable levels of neurodevelopmental disorders in humans, supporting the importance of the NatA complex in normal human development.

Scikit-ribo: Accurate Estimation and Robust Modeling of Translation Dynamics at Codon Resolution

This work was done in collaboration with H. Fang, Y-F. Huang, A. Radhakrishnan, A. Siepel, M. Schatz (CSHL).

Ribosome profiling (Ribo-seq) is a powerful technique for measuring protein translation; however, sampling

errors and biological biases are prevalent and poorly understood. Addressing these issues, we developed Scikit-ribo (https://github.com/schatzlab/scikit-ribo), the first open-source software for accurate genome-wide A-site prediction and translation efficiency (TE) estimation from Ribo-seq and RNA-sequencing (RNA-seq) data. Scikit-ribo accurately identifies A-site locations and reproduces codon elongation rates using several digestion protocols. We showed that commonly used RPKM-derived TE estimation is prone to biases, especially for low-abundance genes. Scikit-ribo introduces a codon-level generalized linear model with ridge penalty that correctly estimates TE while accommodating variable codon elongation rates and mRNA secondary structure. This corrects the TE errors for more than 2000 genes in *Saccharomyces cerevisiae*, which we validated using mass spectrometry of protein abundances, and allows the determination of the Kozak-like sequence directly from Ribo-seq. We also completed an analysis of coverage requirements needed for robust codon-level analysis and quantified the artifacts that can occur from cycloheximide treatment.

Outlier Gene Expression Reveals Recurrent Dysregulation in Rare Disease Pedigrees

M. Doerfel, J. Crain [in collaboration with S. Ballouz, M. Crow, J. Gillis, CSHL]

In disease expression analysis, looking for shared functional signals from a set of genes that show differential expression is commonplace. We examine the complement as a possibility, that disease genes display "outlier" or unexpected expression relative to broader patterns of functional expression variation. Using six families from the rare *TAF1* syndrome disease cohort, we performed family-specific differential expression analyses and found that functional characterization of top candidates enriches for common pathways unlikely to be specifically linked to disease. However, by filtering away common expression changes using known coexpression, we lose all functional enrichment and are left with a small number of outliers characteristic of each proband. Two of these outlier genes are highly recurrent across pedigrees (FDR < 2.63e–05) and are the primary commonality among the cohort as a whole. This suggests that systems analysis may be relevant to rare diseases—principally as a means of filtering out biological signals unrelated to disease.

We have also continued to work on X-linked dystonia-parkinsonism (XDP), which is an adult-onset neurodegenerative disease endemic to the island of Panay, Philippines, where its prevalence is reported to be 5.74 cases per 100,000 individuals with a mean age at onset of 39.7 years. The documented clinical phenotype most frequently combines features of dystonia and parkinsonism in a characteristic temporal progression, beginning with hyperkinetic symptoms at early stages and progressing to predominantly hypokinetic movements at later stages. Various genetic studies have raised the possibility that a defect in *TAF1* may somehow underlie XDP pathogenesis. *TAF1* encodes TATA-binding protein (TBP)-associated factor-1 (*TAF1*), a component of the TFIID complex that mediates transcription by RNA polymerase II (RNAPII) and has emerged in recent years as a significant disease target. In addition to the XDP-related sequence variants in this gene, we reported coding variation in *TAF1* to be associated with severe neurodevelopmental defects and intellectual disability. Given the essential function of *TAF1* in transcriptional regulation in all cells, it is not known how these reported XDP-specific sequence variants may cause tissue-specific defects and/or highly specific clinical phenotypes. Given the apparently distinct phenotypic spectra between XDP, which we propose to be associated with a noncoding mechanism that results in partial reduction of *TAF1* in males, and the recently reported *TAF1* neurodevelopmental syndrome, we compared phenotypic data available between individuals in both cohorts. Videotaped clinical examinations from five individuals from three of these families were evaluated by the movement disorder neurologists who characterized the clinical phenotypes in our XDP cohort. Two affected individuals (IV-3, age 21 years, and III-2, age 23 years) from the previously described extended family with a large copy number variant (CNV) duplication involving *TAF1* had severe neurologic deterioration, including prominent oropharyngeal dysphagia, which is characteristic of XDP. They also displayed cervical and jaw-opening dystonias, which are cardinal features of XDP. The probands in the other families showed milder dystonic features, but two of these individuals (brothers from Family 1) were much younger in age (12 and 14 years old). We also explored the expression data from the XDP fibroblast and neural derivative cell models and from RNA isolated directly from blood from six TAF1 neurodevelopmental syndrome families, which

we previously reported, but detected no statistically significant overlap between DEGs among the different cohorts and cell types.

Expanding Collection and Sequencing of Other Rare Genetic Syndromes

Y. Wu [in collaboration with H. Fang, CSHL; R. Robison, Utah; K. Wang, California; D. Goldstein, Columbia; A. Rope, Oregon; J. Lupski and J. Posey, Baylor; and others]

We continue to meet and collect many families in Utah and elsewhere with very rare, idiopathic genetic syndromes, which we are sequencing and analyzing with many groups. The total number of DNA samples collected to date is approaching 2000, and this includes detailed phenotyping information. We have been making extensive use of Human Phenotype Ontology terms, and the PI was an author of a review concerning the current progress with the development and integration of HPO in various research settings. The PI continues to publish case reports and cohort studies of various rare genetic syndromes.

Collaborating on Genetics of Tourette Syndrome

This work was done in collaboration with the Tourette Syndrome Association International Consortium for Genetics.

The PI continues to collaborate on this international effort to understand the genetics of Tourette syndrome (TS). Psychiatric comorbidity is common in TS; when present, these conditions typically cause more distress and impairment than do tics. High rates of attention-deficit/hyperactivity disorder (ADHD) and obsessive-compulsive disorder (OCD) are well documented and thought to be core components of the TS phenotype; however, few studies have fully characterized other comorbidities. We therefore continue to characterize the prevalence and impact of psychiatric comorbidity in a large sample of individuals with TS and their family members.

PUBLICATIONS

Darrow SM, Grados M, Sandor P, Hirschtritt ME, Illmann C, Osiecki L, Dion Y, King R, Pauls D, Budman CL, et al. 2017. Autism spectrum symptoms in a Tourette's disorder sample. *J Am Acad Child Adolesc Psychiatry* **56:** 610–617.

Darrow SM, Hirschtritt ME, Davis LK, Illmann C, Osiecki L, Grados M, Sandor P, Dion Y, King R, Pauls D, et al. 2017. Identification of two heritable cross-disorder endophenotypes for Tourette syndrome. *Am J Psychiatry* **174:** 387–396.

Dorfel MJ, Fang H, Crain J, Klingener M, Weiser J, Lyon GJ. 2017. Proteomic and genomic characterization of a yeast model for Ogden syndrome. *Yeast* **34:** 19–37.

Fang H, Wu Y, Yang H, Yoon M, Jimenez-Barron LT, Mittelman D, Robison R, Wang K, Lyon GJ. 2017. Whole genome sequencing of one complex pedigree illustrates challenges with genomic medicine. *BMC Med Genomics* **10:** 10.

Huang AY, Yu D, Davis LK, Sul JH, Tsetsos F, Ramensky V, Zelaya I, Ramos EM, Osiecki L, Chen JA, et al. 2017. Rare copy number variants in *NRXN1* and *CNTN6* increase risk for Tourette syndrome. *Neuron* **94:** 1101–1111.

Kohler S, Vasilevsky NA, Engelstad M, Foster E, McMurry J, Ayme S, Baynam G, Bello SM, Boerkoel CF, Boycott KM, et al. 2017. The Human Phenotype Ontology in 2017**.** *Nucleic Acids Res* **45:** D865–D876.

In Press

Fang H, Huang YF, Radhakrishnan A, Siepel A, Lyon GJ, Schatz MC. 2018. Scikit-ribo enables accurate estimation and robust modeling of translation dynamics at codon resolution. *Cell Syst* **6:** 180–191.e4.

Greenberg E, Tung ES, Gauvin C, Osiecki L, Yang KG, Curley E, Essa A, Illmann C, Sandor P, Dion Y, et al. 2018. Prevalence and predictors of hair pulling disorder and excoriation disorder in Tourette syndrome. *Eur Child Adolesc Psychiatry* **27:** 569–579.

Hirschtritt ME, Darrow SM, Illmann C, Osiecki L, Grados M, Sandor P, Dion Y, King RA, Pauls D, Budman CL, et al. 2018. Genetic and phenotypic overlap of specific obsessive–compulsive and attention-deficit/hyperactive subtypes with Tourette syndrome. *Psychol Med* **48:** 279–293.

ANALYSES OF GENOME STRUCTURE AND FUNCTION

W.R. McCombie Y. Berstein, B. Dong, E. Ghiban, S. Goodwin, S. Iyer, M. Kramer, J. Lihm, S. Mavruk Eskipehlivan, O. Mendivil Ramos, S. Muller, A. Qui, S. Teplin, R. Wappel

In 2017, we focused more of our efforts on long-read sequencing in several contexts. These included endeavors ranging from plant genomics to characterization of cancer cell lines and organoids. Long-read sequencing gives dramatically higher resolution of complex regions than the standard short-read sequencers that have been available for about 10 years. We have also continued probing the genetics and epigenetics of psychiatric diseases in 2017.

Application of Long-Read Technologies to Probe the Genomic Architecture of Tumor Organoids

S. Goodwin, J. Lihm, M. Kramer, R. Wappel, W.R. McCombie [in collaboration with G. Arun and D. Spector, CSHL; K. Kodtroff, Northwell Health; M. Schatz, R. Sherman, I. Lee, and W. Timp, Johns Hopkins University; F. Sedlazeck, Baylor College of Medicine]

Cancers are frequently characterized by structural rearrangements that can confer novel phenotypes, including drug resistance, to the cancerous tissue. There is a need to accurately characterize the repertoire of structural event changes within individual cancer genomes to provide the most complete insight into tumor pathology. The development of "next-generation" DNA sequencing based on short sequence reads, typically 150 bases or less, has revolutionized the field of individual cancer genomics. However, recent advances in long-read sequencing have shown that short-read approaches can miss a significant number of structural variants. Unfortunately, the amount of material derived from biopsies is insufficient for comprehensive, long-read sequencing. However, organoids, a three-dimensional cell culture derived from a primary tumor sample, can be grown in the lab and provide sufficient material for long-read sequencing. Long-read technology, along with short-read data, can be used to deeply probe the genomic architecture of organoids to more fully understand the events driving carcinogenesis in that particular tumor, as well as to evaluate the genomic stability of tumors in organoid culture.

The Oxford Nanopore PromethION offers a unique opportunity to explore genome architecture with long reads at low cost. This technology also offers the opportunity to explore the methylation profile of DNA without the need for additional DNA preparation or manipulation; information about the methylation state of the DNA can be derived from the raw sequencing data. This project aims to perform high-coverage (>60×), long-read sequencing of tumor-normal organoid pairs. Both Oxford Nanopore and PacBio sequencing will be performed to identify differences in the performance of each technology. The methylation state of the organoids will be measured using Oxford Nanopore and Illumina methyl-sequencing (methyl-seq). Illumina RNA-Seq will be used to detect fusion transcripts that may derive from DNA fusion events.

In 2017, two breast cancer organoids were obtained from the Spector group at CSHL. An additional cancer cell line, SKBR3, was also sequenced on the Oxford Nanopore instrument so comparisons with previously generated PacBio data could be performed. Preliminary analysis of structural events and the methylation state was executed using the available 30× Oxford Nanopore coverage for SKBR3 and the tumor organoid. At this time, the normal organoid has not generated enough DNA for high-depth sequencing.

Comparisons of the Oxford Nanopore and PacBio data for the organoid and SKBR3 samples have shown a high concordance among the structural variant calls. More than 15,000 structural variants were identified in the organoid sample. Of these variants, more than 10,000 are shared between Oxford Nanopore and PacBio, whereas around 3000 cells were unique to Oxford Nanopore and 5000 were unique to PacBio. These results are quite striking when compared with

the Illumina data. The Illumina data only identified 6000 variants, 2000 of which were shared with Oxford Nanopore Technologies (ONT) and Pacific Biosciences (PacBio). The difference in the number of unique variants found in each technology is attributable to several factors: Variants uniquely found in PacBio/ONT are likely real variants that are undetectable in short-read technology, whereas unique variants in the Illumina data are more likely to be artifacts. The unique variants found between ONT and PacBio are likely because of the different error profiles found in each technology, as well as the fact that there is deeper PacBio coverage. Increasing the Oxford Nanopore coverage to more than 60 will likely improve the concordance between the technologies.

The Timp group at Johns Hopkins performed methylation analysis on the Oxford Nanopore data. Both SKBR3 and the tumor organoid showed similar patterns of hyper- and hypomethylation, indicating cancer-specific methylation patterns. Patterns of hyper- and hypomethylation have been identified in the promoter regions of several genes; however, RNA-Seq has not yet been performed to determine whether this has an effect on expression. There also appears to be some evidence of hypomethylation flanking deletion sites in SKBR3 and the tumor organoid, implying a relationship between cancer-specific methylation patterns and structural variation. Additional work will need to be performed to further explore these findings.

Assessment of Tumorigenicity in Growth of Organoid Cancer Models

W.R. McCombie, S. Goodwin, M. Kramer, S. Muller [in collaboration with D. Tuveson, D. Spector, P. Sridevi, D. Plenker, G. Arun, and S. Bhatia, CSHL; V. Corbo, University of Verona]

Recent studies have shown that organoids can be superior to cell lines for modeling cancer treatment response. However, organoid culture can be complex, and the presence of normal cells may adulterate the purity of the model. In collaboration with the Tuveson and Spector labs, we have developed a targeted strategy to sequence and analyze cancer genes from organoid models grown at CSHL to confirm that each model is tumor derived. The sequencing panel was designed using NimbleGen V3 probes to capture 140 genes known to be involved in cancer, along with probes to detect mouse and mycoplasma contamination. Organoid DNA was sheared and barcoded and the resulting libraries were then pooled in sets of 24–48. After hybridization with the probes, the samples were sequenced on Illumina NextSeq instruments to achieve >400× coverage. This depth allows for more sensitive detection of low-frequency somatic variants. We then processed the data with a pipeline consisting of BWA alignment, processing with SAMtools, removal of polymerase chain reaction (PCR) duplicates with Picard, and filtering with BamTools for mapping quality and proper read pairing. Variants were called using VarScan2, and annotated with ANNOVAR. We then selected rare loss-of-function variants (nonsense, frameshift, splicing) with frequency <1% in the gnomAD, ExAC, EVS, and 1000 Genomes databases. Missense and in-frame indel variants were selected if they were noted as pathogenic by ClinVar, or were both rare and annotated as pathogenic by COSMIC, or found to be present in the TCGA cohorts. Oncoplots were then generated from these candidate variants using maftools.

Our initial test captures were performed across models from several cancer types (lung, breast, pancreas, and colon). This preliminary work showed that we could detect driver mutations in nine models and those mutations had high concordance with matched primary tumor samples. The models that did not show clear drivers were categorized as provisionally derived from normal tissue, and additional testing is planned to identify copy number variants that may be indicative of tumor types that do not harbor drivers in our capture gene set. These models may be used as a resource for cancer research in collaboration with Leidos and NCI.

Definition of the Wheat Epigenome

W.R. McCombie, S. Goodwin, E. Ghiban [in collaboration with R. Martienssen, M. Regulski, J. Han, C. LeBlanc, and M. Donoghue, CSHL; L. Gardiner, Earlham Institute; A. Hall and J. Kenny, University of Liverpool; K. Mayer and M. Pfeifer, Helmholtz-Zentrum Muenchen; M. Bevan, John Innes Centre Institute]

The ERA-CAPS project is to define the epigenome of elite wheat strains. The first steps in this project will comprise high-throughput sequencing of bisulfite-converted DNA. CSHL will develop and optimize pipelines for this portion of the project. Initially, 50× coverage of ~1000-bp paired-end sequencing

data was planned for CS42. However, because of work being performed at other institutions and difficulties with sequencing performance of CS42, it was decided in late 2017 to switch to the Paragon wheat variety for the epigenome profiling. Multiple bisulfite libraries of Paragon were prepared from the sample tissue and tested for performance on the MiSeq. Libraries with a lower PCR duplicate rate will be sequenced on the HiSeq 2500 in paired-end 250-bp, rapid mode. This strategy will ensure a lower PCR duplication rate overall. High coverage (>50×) bisulfite sequencing of the Paragon wheat strain will be completed in 2018. Transcriptome sequencing and small RNA sequencing (RNA-Seq) of CS42, Paragon, and other strains will also be performed. These steps will correlate gene expression with epigenetic state and explore transposon methylation within the genome. Following these initial steps and capture probe, exome panel for methylation analysis will be designed. The panel will allow for the capture and high-depth (~100×) sequencing of the exome and flanking sequences (~24 Mb) without the loss of the methylation state. Preliminary analysis and method development will lead to several additional studies, including analysis of the epigenome of hybrid progenitors and their germline, and exploring the relationship between epigenetic modification and the environment. Comparison of the epigenome work to tissue-specific transcriptome work will also be performed in 2018.

The MaizeCODE Project

W.R. McCombie, S. Goodwin, M. Kramer, E. Ghiban [in collaboration with T. Gingeras, C. Danyko, D. Jackson, R. Martienssen, M. Regulski, D. Micklos, M. Schatz, and D. Ware, CSHL; Ken Birnbaum, NYU; Hank Bass, Florida State]

MaizeCODE is a consortium effort to explore maize varieties at the genomic, transcriptomic, and epigenetic levels. Building on the efforts of our previous report, we have completed the sequencing of the NC350 genome using PacBio Sequel long reads, achieving 55× coverage with an average read length of >7 kb. Combining the long-read data with 10× Chromium-linked Illumina reads allowed our collaborators in the Schatz lab to produce a high-quality assembly with a scaffold N50 of >1.2 Mb, comparable to the published B73 genome. In addition, we have completed sequencing of multiple RNA libraries produced by the Gingeras lab, including NC350 long RNA-Seq and RAMPAGE, as well as ears, meristem, and root tip fluorescence-activated cell-sorted RNA-Seq.

All data is available at the NCBI SRA and on CyVerse Data Commons. Future work will include sequencing the long RNA-Seq and RAMPAGE libraries of the W22 and B73 strains, sequencing short RNA libraries for all strains, and genome sequencing of the maize progenitor Teosinte, using long reads generated by the new Oxford Nanopore PromethION instrument.

Improved Assembly of the B73 Maize Genome

S. Goodwin, S. Mavruk Eskipehlivan, W.R. McCombie [in collaboration with D. Ware, Y. Jiao, B. Wang, M. Campbell, J. Stein, X. Wei, M. Regulski, S. Kumari, and A. Olson, CSHL; E. Antoniou, BioReference Laboratories; P. Peluso, C. Chin, and D. Rank, Pacific Biosciences; J. Shi, T. Liang, and A. Hastie, BioNano Genomics; M. Stitzer, M. May, and J. Ross-Ibarra, University of California Davis; K. Guille and M. McMullen, USDA-ARS Plant Genomics Research Unit; J. Gent and R.K. Dawe, University of Georgia; K. Schneider, T. Wolfgruber, and G. Presting, University of Hawai'i; N. Springer, University of Minnesota]

Complete and accurate genome annotations can only be achieved when referencing reference genomes of the highest possible quality. These resources facilitate research into biological processes that are critical in trait determination, which, in turn, are critical for developing sustainable agricultural practices. Next-generation sequencing has been instrumental in the creation of many reference genomes for plant species; however, limitations of the technologies used have resulted in references that are fragmented and missing complex repeat regions. The use of long-read, single-molecule sequencing technology provides contiguous sequencing that can be used to notably increase the completeness and accuracy of plant genomes. To highlight the power of this technology and provide a highly accurate reference genome for maize, the B73 maize strain was sequenced to more than 60× coverage on the Pacific Biosciences RS II instrument. An additional 60× coverage was generated via the BioNano optical mapping instrument.

The assembly of the PacBio data consisted of 2958 contigs in which half of the total assembly was made up of contigs of >1.2 Mb. This assembly was then integrated with high-quality BioNano optical mapping

and previously generated bacterial artificial chromosome (BAC) assembly data. The new reference assembly has a 240-fold higher contiguity than the recently published short-read assembly of the PH207 maize cultivar (contig N50, 1180 kb vs. 5 kb) (Hirsch 2016). This assembly also showed a 53-fold improvement in contiguity over the BAC clone-based assembly with capillary-based sequencing of B73 (contig N50, 40 kb) (Schnable 2006). These data were analyzed by the Ware group in 2017. The results are published in *Nature Letters* as of June 2018. These data support the hypothesis that high-quality, long-read sequencing is essential for complete and accurate annotation of plant genomes. This additionally shows that there are arguments to be made for strain-specific assembly to detect many of the structural variants present in this highly diverse species.

Sequencing and Assembly of Tomato Genomes with Oxford Nanopore Technology

S. Goodwin, R. Wappel, W.R. McCombie [in collaboration with S. Soyk, X. Wang, Z. Lemmon, and Z. Lippman, CSHL; F. Sedlazeck, Baylor College of Medicine; M. Alonge and M. Schatz, Johns Hopkins University]

Tomatoes are important fruit-bearing, and thus flower-producing, plants. Solanacea is a diverse family containing both species that produce one flower per inflorescence (such as pepper and petunia) and species that produce hundreds of flowers (such as some varieties of wild tomato). Study of tomato and related species can reveal the genetic and molecular mechanisms that drive inflorescence development and diversity. Although many tomato genomes have been sequenced via short-read sequencing technology, these methods miss many important structural events that may drive phenotype. Long-read technologies such as Oxford Nanopore can resolve these variations at very low cost. In 2017, the McCombie lab, in collaboration with the Lippman lab, began a project to sequence a large number of tomato genomes. 40× coverage of the M82 tomato cultivar was generated using the Oxford Nanopore MinION and GridION instruments. The average read lengths were >10 kb with >5 Gb generated per cell. This performance is comparable with PacBio Sequel technology at a lower cost. This work will be continued into 2018 with the goal of sequencing 100 tomato cultivars.

Sequencing, Assembly, and Annotation of the Flame Seedless Grape Variety

S. Goodwin, R. Wappel, S. Mavruk Eskipehlivan, W.R. McCombie [in collaboration with M. Campbell, M. Regulski, and D. Ware, CSHL; L. Cadle-Davidson, C. Ledbetter, and R. Naegele, USDA; S. Qi and F. Gouker, Cornell University]

Grapes are an important agricultural product grown around the world for food, juice, wine, and oil production. The "Flame seedless" grape variety is one of the most widely planted table grape varieties. Currently there is no high-quality, long-read assembly available for this plant. A contiguous assembly of this cultivar will provide insight into traits that differentiate seedless table grapes from wine grapes and reveal important structural elements that may contribute to disease resistance. This project began in late 2016 with the aim of sequencing the Flame seedless variety on the PacBio RS II. Initially, sequencing performance varied significantly for this sample, with cell yield ranging from 100 Mb to >900 Mb and average read lengths ranging from 2 kb to 9 kb. It is believed that an adjunct, sugar perhaps, was negatively impacting sequencing performance. In 2017, Swift Biosciences released a new kit for PacBio sequencing that did not rely on the SMRTbell technology, which allowed library prep to be performed in a single day and with fewer pipetting steps, thereby improving sequencing performance. This new method increased flow cell yield to 1 Gb and increased read length from 9711 bp to 12,289 bp, whereas the proportion of reads of >10 kb increased from 15% to 39.4%. By the end of 2017, >40 Gb of data were generated for the 50-Gb target. In addition, flame grape material, along with cintera, another grape variety, was sent to 10× Genomics for synthetic long-read sequencing. The PacBio sequencing for this project will continue into 2018 when work on assembly and annotation will begin.

The Genetic Basis of Ethnic Disparity in Colon Cancer

M. Kramer, S. Goodwin, R. Wappel, S. Muller, W.R. McCombie [in collaboration with X. Wang, X. Yu, E. Li, and J. Williams, Stony Brook University]

In recent years, studies have shown differences in the mutational landscape of African American (AA) colon cancers compared with those of Caucasian Americans

(CAs), which may shed light on the higher incidence and mortality rate of the disease in AAs. To address the challenges we previously reported in our collaboration to produce linked DNA, RNA, and methylation profiles for underrepresented minorities, we performed exome capture on DNA from new extraction procedures from 16 tumor-normal pairs (including 10 AA samples) and were able to achieve >80% of bases at ≥30× coverage depth for almost all samples. We were able to detect loss-of-function mutations in genes associated with colon cancer, including APC and TP53 as well as activating mutations in KRAS. Interestingly, even within the small subset of AA patients tested, we identified an intronic germline susceptibility variant found at higher frequency in AAs (Wang et al. 2017) in one of the downstate samples. In another AA sample, we found a missense variant that is predicted to be deleterious in the *FLCN* gene, which was identified as a driver gene in AA colon cancers by Guda et al. in 2015. We also resequenced 13 tumor-normal pairs with an optimized reduced representation bisulfite sequencing (RRBS) protocol, which showed improved MspI digestion.

These data are being analyzed with the previously completed RNA sequencing of 24 samples (including 10 tumor-normal pairs). In 2018, we will expand analysis of the somatic and germline variants and further explore correlations between these mutations and changes in expression.

Whole-Genome Sequence Analysis of the Original DISC1 Pedigree

J. Lihm, M. Kramer, W.R. McCombie [in collaboration with D. Blackwood, D. Porteous, P. Thomson, N. Ryan, and I. Deary, University of Edinburgh; A. McRae, University of Queensland]

We had previously described our work on the targeted sequencing of genes that interact with and are regulated by DISC1 in more than 1500 cases and controls. We found an association between rare disruptive variants in the regulome and schizophrenia, as well as a negative effect on cognitive ability when such variants perturb the interactome. In 2017, our study was accepted and published in *Molecular Psychiatry* (Teng et al.). Continuing our collaboration with the Porteous Lab on the whole-genome sequencing (WGS) of the extended DISC1 pedigree, this year we focused on the detection of large structural variations, which are highly shared among affected members. We called copy number variations (CNVs) on each of the 47 whole-genome samples in the DISC1 pedigree, fine-tuning the segmentation algorithm from Yoon et al. (*Genome Res 19:* 1586 [2009]) to our data. We computed read-depth signals with high mapping quality reads (>20). Read-depth signals were controlled for background variability using 14 WGS in-house samples that were not diagnosed with any psychiatric disorders. Then, we applied the segmentation algorithm to the normalized signals to call deletion and duplication regions. Deletion regions showed clear patterns as either heterozygous deletions with around half the average of genome-wide signals or homozygous deletions with nearly zero signals. Duplication signals were more ambiguous; thus we focused on the analysis of deleted regions. We first filtered the initial calls by size (>1 Kbp) and the amplitude of normalized signals, and selected the regions that were preferentially shared between affected members of each nuclear family. After visual investigation, 27 unique deletion regions were selected with sizes from 1.2 kb to 145 kb, and two of them were located within nine linkage regions found in this pedigree. One was located in an intronic region of *CNTN5* (*chr11*) and the other was in the intergenic region between *FSTL5* and *MIR4454* (*chr4*). The two deletion regions were validated by PCR amplification and Sanger sequencing. These results were submitted for publication to *Molecular Psychiatry* in 2017.

The Epigenome of Fear

J. Lihm, W.R. McCombie [in collaboration with S. Ahrens, B. Li, and J. Gillis, CSHL; S. McCarthy, Regeneron]

Previously we studied the sensitivity of ATAC-seq peaks by downsampling reads. In 2017, we continued to strengthen the analysis of the 193 samples downloaded from 12 studies for general profiling of genome-wide ATAC-seq signals and the relation of chromatin accessibility to gene expression. First, we found that the signals within TSS (transcription start site) regions (±1 Kbp of TSS) were more robust to the change in read coverage compared with other regulatory regions. We quantified the number of genes with peaks in TSS regions, and called TSS-accessible genes, versus the number of reads. As the number of reads increased, the number of TSS-accessible genes grew larger as well

and converged to 12,000 genes at 30 M reads, whereas the number of peaks did not show the same pattern of convergence. This converging pattern suggested a quality control criterion on the number of reads to be >30 M reads. Among the 193 collected samples, 37 (19.2%) samples met this criterion. The utility of this criterion is also confirmed by the fold change of ATAC-seq read signals within peaks versus the background signals. We defined poor-quality peaks as the ones with fold change of <2. The fraction of poor-quality peaks dropped exponentially as the number of reads increased. Around 90% of peaks passed the fold change criterion for samples that passed our quality control (QC) criterion as having >30 M reads.

Also, we found that ATAC-seq signals are more robust in TSS regions than in other genomic regions. We used downsampling analysis to confirm this observation. We sampled 90% of the original reads and called peaks with the sampled data. After peak calling in both full data and 90% of the data, we compared how many peaks were gained and lost. The results showed that on average 10.0% of peaks were not called genome-wide in the sampled data ranging from 5.3% to 19.4%. When we assessed peaks in TSS regions, however, only 4.4% of TSS-accessible genes were lost in the sampled data on average. This observation suggested that ATAC-seq signals in TSS regions were less impacted by the loss or change of underlying data. Thus, we further focused on the gene-level analysis, looking at peaks and read signals around TSS regions.

We next studied genes that are commonly accessible through all the samples. There was a total of 451 genes that were called in all 193 samples. The 451 commonly accessible genes could be either functionally constitutive or technically prone to have reads. We measured the average expression level of these genes from a large collection of gene expression studies, using the Gemma database. We calculated the rank of mean expression of each gene across different experiments, scaled from 0 (lowest) to 1 (highest). On average, the 438 genes that were found in the Gemma database were in the top quartile rank (mean expression rank 0.75), whereas the variations of expression level were relatively lower (mean SD, rank 0.34). In other words, these 438 genes were relatively highly expressed in a consistent manner across experiments, suggesting these genes could be putative housekeeping genes.

Last, we applied the results from the meta-analysis to our mouse brain data (18 samples in total). We dropped two samples that did not meet our QC criterion and re-analyzed the remaining 16 samples. We continued to profile robust accessible regions in the mouse brain by sampling 10% of reads without replacement 10 times to use the entire set of reads. We selected genes that were called in nine or more replicates out of 10 and detected 6.3-K TSS-accessible genes per sample on average. Despite the relatively small number of genes, >94% of 451 common genes were included, confirming the strong signals of the commonly accessible genes.

In 2018, we will continue to analyze differential chromatin accessibility of mouse brain regions between amygdala and cortex, fear conditioning versus control, and mutant versus wild-type mice. We are going to construct a unified analysis pipeline that can be generally applicable to any ATAC-seq mouse data. We will also focus on refining the statistical method for detecting differentially accessible regions between two conditions.

PUBLICATIONS

Daneshjou R, Wang Y, Bromberg Y, Bovo S, Martelli PL, Babbi G, Lena PD, Casadio R, Edwards M, Gifford D, et al. 2017. Working towards precision medicine: Predicting phenotypes from exomes in the Critical Assessment of Genome Interpretation (CAGI) challenges. *Hum Mutat* **38:** 1182–1192.

Goodwin S, Wappel R, McCombie WR. 2017. 1D genome sequencing on the Oxford Nanopore MinION. *Curr Protoc Hum Genet* **94:** 18.11.1–18.11.14.

Jiao Y, Peluso P, Shi J, Liang T, Stitzer MC, Wang B, Campbell M, Stein JC, Wei X, Chin C-S, et al. 2017. Improved maize reference genome with single molecule technologies. *Nature* **546:** 524–527.

Miller JR, Zhou P, Mudge J, Gurtowski J, Lee H, Ramaraj T, Walenz BP, Liu J, Stupar RM, Denny R, et al. 2017. Hybrid assembly with long and short reads improves discovery of gene family expansions. *BMC Genomics* **18:** 541.

Monson ET, Pirooznia M, Parla J, Kramer M, Goes FS, Gaine ME, Gaynor SC, de Klerk K, Jancic D, Karchin R, et al. 2017. Assessment of whole-exome sequence data in attempted suicide within a bipolar disorder cohort. *Mol Neuropsychiatry* **3:** 1–11.

Stanley R, Martin P-M, Ross A, Freitas A, Iafrati J, Moyer C, Brumback A, Pirooznia M, McCombie WR, Potash J, et al. 2017. Wnt/β-catenin pathway contributions to dendritic spine and glutamatergic synapse formation responsive to lithium-mediated GSK3 inhibition. *Biol Psychiatry* **81:** S286.

In Press

Teng S, Thomson PA, McCarthy S, Kramer M, Muller S, Lihm J, Morris S, Soares DC, Hennah W, Harris S, et al. 2018. Rare disruptive variants in the DISC1 Interactome and Regulome: Association with cognitive ability and schizophrenia. *Mol Psychiatry* **23:** 1270–1277.

PLANT GENOME ORGANIZATION, EVOLUTION, AND FUNCTION

D. Ware
M.S. Campbell
K. Chougule
N. Gladman
F. Hu
Y. Jiao
V. Kumar
S. Kumari
Y.K. Lee
M.N.D.S. Leite
Z. Lu
D. Muna
A. Olson
M. Regulski
J. Stein
M.K. Tello-Ruiz
J. Thomason
P. Van Buren
B. Wang
L. Wang
X. Wang
S. Wei
L. Zhang

The Ware lab has two primary goals: (1) understanding plant genome function, agriculturally important crop plants; and (2) development of tools, data sources, and resources for the genomics research community.

The Ware lab is a dynamic research group and its composition has continued to evolve, keeping pace with its research objectives. During this period, one of our postdoctoral researchers, Michael Campbell, completed his research program and went onto newer career opportunities, and Mariana Neves Dos Santos Leite left as a research technician. During this period, Nicholas Gladman joined the team as postdoctoral researcher, Demitri Muna as bioinformatics analyst, and Fangle Hu as research technician. In addition, George Wang and Pragati Muthukumar completed their Undergraduate Research Program (URP) and internship program, respectively, during this year.

Plant Genome Research

In the last decade, the sequencing and annotation of complete plant genomes has helped us understand plant function and evolution, as well as how to alter economically important traits. Efforts in many disparate disciplines are required to generate reference genomes. The work at the Ware lab often starts with laboratory scientists who use wet chemistry to generate the raw sequence data. Next, computational biologists and bioinformaticians kick off a series of computational steps to interpret the raw data. The process of interpretation involves the assembly of raw sequence reads into overlapping segments ("contigs"), which are combined to create a scaffold. This scaffold, in turn, discerns the position, relative order, and orientation of contigs within the chromosomes. The next step is annotation, the discovery and description of genes and other functional elements, and homologies (evolutionary relationships) with other genomes. This information must be faithfully communicated and visualized in web-based platforms such as Gramene.

All of these activities are rapidly evolving with the advent of new sequencing technologies, algorithms, and data-handling requirements. For example, high-depth and low-cost sequencing of RNA transcripts is providing a vast stream of new evidence that informs genome annotation; this, in turn, has spurred the development of new software for modeling and performing genome annotation. Low-cost sequencing has also made it possible to ask whole new classes of questions, moving beyond the generation of single references for individual species and supporting the development of multispecies representation as a "pangenome." Ongoing projects within the maize, rice, *Arabidopsis*, sorghum, and grape research communities are now sequencing hundreds or thousands of genotypic backgrounds, chosen from carefully constructed populations, wild populations, and breeding germplasms in each species. Information about genetic variation is helping scientists to understand the genetic basis of phenotypic traits and the origins of domesticated crop species. New technologies are also driving research in epigenetics, the study of heritable variation not attributable to changes in the underlying genome sequence. Epigenetic mechanisms include modification of DNA by methylation and various forms of histone modification, which can cause changes in gene expression and other phenotypes. Both types of modifications can be studied using new sequencing technologies and analytical methods.

Updating the Maize B73 Reference Genome and Annotation Using Long Sequence Read Technologies

Y. Jiao, B. Wang, M. Regulski, A. Olson, J. Stein, S. Kumari, M. Campbell [in collaboration with The Maize Genome Sequencing Consortium; R. McCombie and S. Goodwin, CSHL; K. Guill and M. McMullen, Missouri University, USDA ARS; J. Ross-Ibarra, UC Davis; K. Dawe, University of Georgia; D. Rank, P. Peluso, and T. Clark, Pacific Biosciences; A. Hastie, BioNano Genomics]

Maize is an important feed and fuel crop, as well as a model system in developmental genetics. A complete and accurate reference genome is imperative for sustained progress in understanding the genetic basis of trait variation and crop improvement. The Ware lab has played a leading role in development and stewardship of the maize reference genome for more than a decade. The 2009 release of the B73 reference sequence was a milestone in plant genomics research because of the unprecedented size and complexity of the maize genome. Through several updates, this foundational resource has remained the principal genome reference for the maize research community. Yet, it continues to be a work in progress, with gaps and misassemblies that have defied available sequencing technologies, especially over the highly repetitive regions that are also the most dynamic and rapidly evolving.

This year marks a new milestone with the release of an entirely new reference assembly, designated B73 RefGenV4. In collaboration with Pacific Biosciences, we sequenced maize to ~65× coverage using single molecule real-time (SMRT) technology. These exceptionally long reads, typically >10 Kb, were able to span not only genes, but also the extensive intergenic and repetitive regions rich in transposons. Using an optimized correction and assembly pipeline, we built a de novo assembly consisting of 2958 contigs and totaling 2.10 Gb, with more than half of this length in contigs of >1.2 Mb. This represents a 52-fold improvement in assembly contiguity compared with the previous bacterial artificial chromosome (BAC)-based maize assembly, whereas nucleotide agreement was maintained at >99.9%. Optical maps of the maize genome, developed at BioNano Genomics, enabled scaffolding of these contigs into the 10 chromosomes of maize, thereby placing >99% of maize genes. Many improvements in gene order and orientation were found. The new sequence fills in many intergenic regions that harbor transposons and regulatory regions influencing gene expression. In addition, comparative optical mapping of two other inbreds revealed a prevalence of deletions in the region of low gene density and maize lineage-specific genes (Jiao et al. 2017).

Uncertainties about the complete structure of mRNA transcripts, particularly with respect to alternatively spliced isoforms, can be a limiting factor for research in the system. In addition to working on the reference genome sequence, we are using the same single-molecule sequencing technology to investigate the maize transcriptome. For this work we have sampled full-length cDNAs from six tissues of the maize inbred line B73. These were barcoded, pooled, size-fractionated (<1 kb, 1–2 kb, 2–3 kb, 3–5 kb, 4–6 kb, and 5–10 kb), and sequenced on the PacBio RS II platform with P6-C4 chemistry. We were able to capture 111,151 transcripts, representing ~70% of the genes annotated in the current maize RefGenV3 genome assembly. A large proportion of transcripts (57%) are novel. We were able to validate ~90% of the transcript splice-site-junctions within high-depth short reads generated from the matched tissues. In addition, we identified a large number of novel long noncoding RNAs (lncRNAs) and fusion transcripts. Our results show that the characterization of the maize B73 transcriptome is far from complete, and maize gene expression is more complex than previously thought (Wang et al., *Nat Commun 7:* 11708 [2016]).

A Sorghum Mutant Resource as an Efficient Platform for Gene Discovery in Grasses

Y. Jiao [in collaboration with Z. Xin and J. Burke, USDA ARS]

Sorghum (*Sorghum bicolor* (L.) Moench) is a versatile C4 crop and a model for research in the family Poaceae. High-quality genome sequence is available for the elite inbred line BTx623, but functional validation of genes remains a challenge because of the limited genomic and germplasm resources available for comprehensive analysis of induced mutations. In this study (Jiao et al., *Plant Cell 28:* 1551 [2016]), we generated 6400 pedigreed M4 mutant pools from ethyl methanesulfonate (EMS)-mutagenized BTx623 seeds through single-seed descent. Whole-genome sequencing of 256 phenotyped mutant lines revealed more than 1.8 million canonical EMS-induced mutations,

affecting >95% of genes in the sorghum genome. The vast majority (97.5%) of the induced mutations were distinct from natural variations. To show the utility of the sequenced sorghum mutant resource, we performed reverse genetics to identify eight genes potentially affecting drought tolerance, three of which had allelic mutations and two of which showed exact co-segregation with the phenotype of interest. Our results establish that a large-scale resource of sequenced pedigreed mutants provides an efficient platform for functional validation of genes in sorghum, thereby accelerating sorghum breeding. Moreover, the findings made in sorghum could be readily translated to other members of the Poaceae via integrated genomics approaches (Jiao et al., *Plant Cell 28:* 1551 [2016]).

Pangenome Annotation

M.S. Campbell [in collaboration with W.H. Majoros, A.S. Allen, and T.E. Reddy, Duke University; C. Holt and M. Yandell, University of Utah; E. DeNardo, Washington University]

The accurate interpretation of genetic variants is critical for characterizing genotype–phenotype associations. Because the effects of genetic variants depend on their local genomic context, accurate genome annotations are essential. Furthermore, as some variants have the potential to disrupt or alter gene structure, variant interpretation efforts stand to gain from the use of individualized annotations that account for differences in gene structure between individuals or strains. We have developed a suite of haplotype-aware software tools for predicting functional changes in gene structure that may result from sequence variants. ACE (assessing changes to exons) converts phased genotype calls to a collection of explicit haplotype sequences, maps transcript annotations onto them, and detects gene-structure changes and their possible repercussions, including several classes of possible loss of function. Using publicly available RNA-sequencing (RNA-Seq) data, we showed that novel transcripts predicted by ACE are commonly supported by spliced RNA-Seq. We also show that ACE predictions confirm earlier results regarding the quantitative effects of nonsense-mediated decay, and that predicted loss-of-function events are highly concordant with patterns of intolerance to mutations across the human population. ACE can be readily applied to diverse species, including animals and plants, making it a broadly suitable tool for use in eukaryotic population-based resequencing projects, particularly for assessing the joint impact of variants at a locus. Future enhancements currently under development include a novel probabilistic model to predict the effects of genetic variants that alter splicing enhancers or create pseudoexons, and quantitative prediction of changes in stoichiometric isoform ratios.

Gramene: Comparative Genomic Resource for Plants

M. Campbell, K. Chougule, Y. Jiao, V. Kumar, S. Kumari, Y.K. Lee, D. Muna, A. Olson, J. Stein, M.K. Tello-Ruiz, B. Wang, S. Wei, L. Zhang [in collaboration with P. Jaiswal, Oregon State University; P. Kersey, I. Papatheodorou, and R. Petryszak, EMBL-European Bioinformatics Institute; L. Stein, Ontario Institute of Cancer Research; C. Taylor, American Society of Plant Biologists; R. Wing, University of Arizona]

The Gramene project provides online reference resources for plant genomes and curated pathways to aid functional genomics research in crops and model plant species. Our website (www.gramene.org) facilitates studies of gene function by combining genome and pathway annotation with experimental data and cross-species comparisons. In other words, the data and tools in Gramene enable plant researchers to use knowledge about gene function in one species to predict gene function in other species. Drawing these connections facilitates translational research in plant development and physiology that influences economically important traits—for example, grain development, flowering time, drought tolerance, and resistance to diseases. In the past year, the project accomplished several major milestones, culminating in our 55th data release (September 2017) since the inception of the project, with a significantly enhanced search user interface and back-end functions. The Gramene team develops comparative genomics databases in collaboration with the Ensembl Plants project at the European Bioinformatics Institute (EMBL-EBI), and collaborates closely with the EBI's Expression Atlas project to provide manually curated, quality-controlled, and analyzed transcriptomic data. We continue to host genome and pathway annotations via: (1) the Ensembl genome browser, and

(2) the Plant Reactome pathways portal. Additions this year to our reference genome resource include one new species: jute (*Corchorus capsularis*). In addition, we made significant updates to the barley (*Hordeum vulgare* V2) reference assembly, and updated the gene annotations for maize (*Zea mays*) and barley (*H. vulgare*). We incorporated new genetic variation data for Japonica rice (QTL data from the Q-TARO database, and SSR/RFLP and QTLs from Gramene's archives remapped to IRGSPv1), *Arabidopsis thaliana* (1001 Genomes Project), and wheat (*Triticum aestivum* EMS mutants). In collaboration with the Earlham Institute scientists, we released genotype data for a mutagenized population of *Triticum aestivum* (Krasileva et al., *Proc Natl Acad Sci 114*: E913 [2017]). Similarly to naturally occurring single-nucleotide polymorphisms (SNPs) and structural variants, the EMS-derived point mutation variants are displayed in the context of gene annotation and inferred functional consequences, which can be assigned to individual accessions within a sampled diversity panel. We added new track hubs, totaling more than 1600 public RNA-Seq studies (nearly 40,000 tracks across 38 plant species), and new automated noncoding RNA (ncRNA) feature alignments across all plant species. Curated baseline gene expression data from the Expression Atlas project (https://www.ebi.ac.uk/gxa/plant/experiments) are now available for 734 experiments in 17 plant species from both our genomes and pathway browsers, as well as a new search interface.

We moved the Gramene homepage to our integrated search interface, whereas blog content and other static pages were moved to news.gramene.org so they could be displayed within the new interface. Gramene's integrated search database and modern user interface leverage diverse annotations to facilitate finding genes through selecting auto-suggested filters with interactive views of the results. The interface guides the user to select filters based on gene identifiers, names, and descriptions, as well as structured annotations such as InterPro domains, Gene Ontology terms, and Plant Reactome pathways. Complex queries can be composed by adding more filters. A paginated list of matching genes is displayed in which each gene can be dynamically loaded to reveal detailed views, which can suggest additional search filters. Each gene match is presented with distinct possible tab views. To the existing location, expression, homology, and X-refs (or cross-references) tab views (Fig. 1), a new Pathway results view was added. We also added a neighborhood conservation mode to the homology tab in the search results. It displays up to 10 protein-coding genes flanking each gene in a gene tree. The genes are color coded based on the gene families represented in the neighborhood of the gene of interest. Each neighborhood links to the Ensembl Synteny browser for the corresponding region. We have also given users the ability to customize the gene tree node labels. Search results are also summarized via facet counting, which can rapidly calculate the genomic distribution of gene search results across all hosted genomes. Our new search API is hosted at data.gramene.org.

We consolidated our dedicated maize genome browser resource (maize4.gramene.org) into the prototype of a maize pangenome browser. Although Gramene's main site (www.gramene.org) is committed to supporting community-recognized annotations for a single reference genome per species, this resource features three maize assemblies: B73 *Zea mays* RefGenV4, *Z. mays* W22, and *Z. mays* PHP and the ancestral teosinte TIL11 assembly. To facilitate comparative and functional genomics research in corn, this resource also hosts a subset of key species, including the dicot model *A. thaliana*, and the monocot model *Oryza sativa Japonica*. Featured annotations will be used to build phylogenetic gene trees and define orthologous and paralogous relationships using the Ensembl Compara gene tree method. These results will yield new insights into the taxonomic origin of genes and patterns of duplication, movement, and loss influenced by genome architecture. The maize V4 assembly and annotations are also available from the main Gramene genome browser.

Pathogen Immunity Genes in Wild Related Species of Rice

J. Stein, K. Chougule, S. Wei [in collaboration with R. Wing, University of Arizona; The International Oryza Map (I-OMAP) Consortium]

Disease pathogens, such as rice blast (*Magnaporthe oryzae*), severely impact rice production and pose an increasing threat as climate change alters the geographical range of pests in the future. Breeding for natural host resistance is a proven strategy

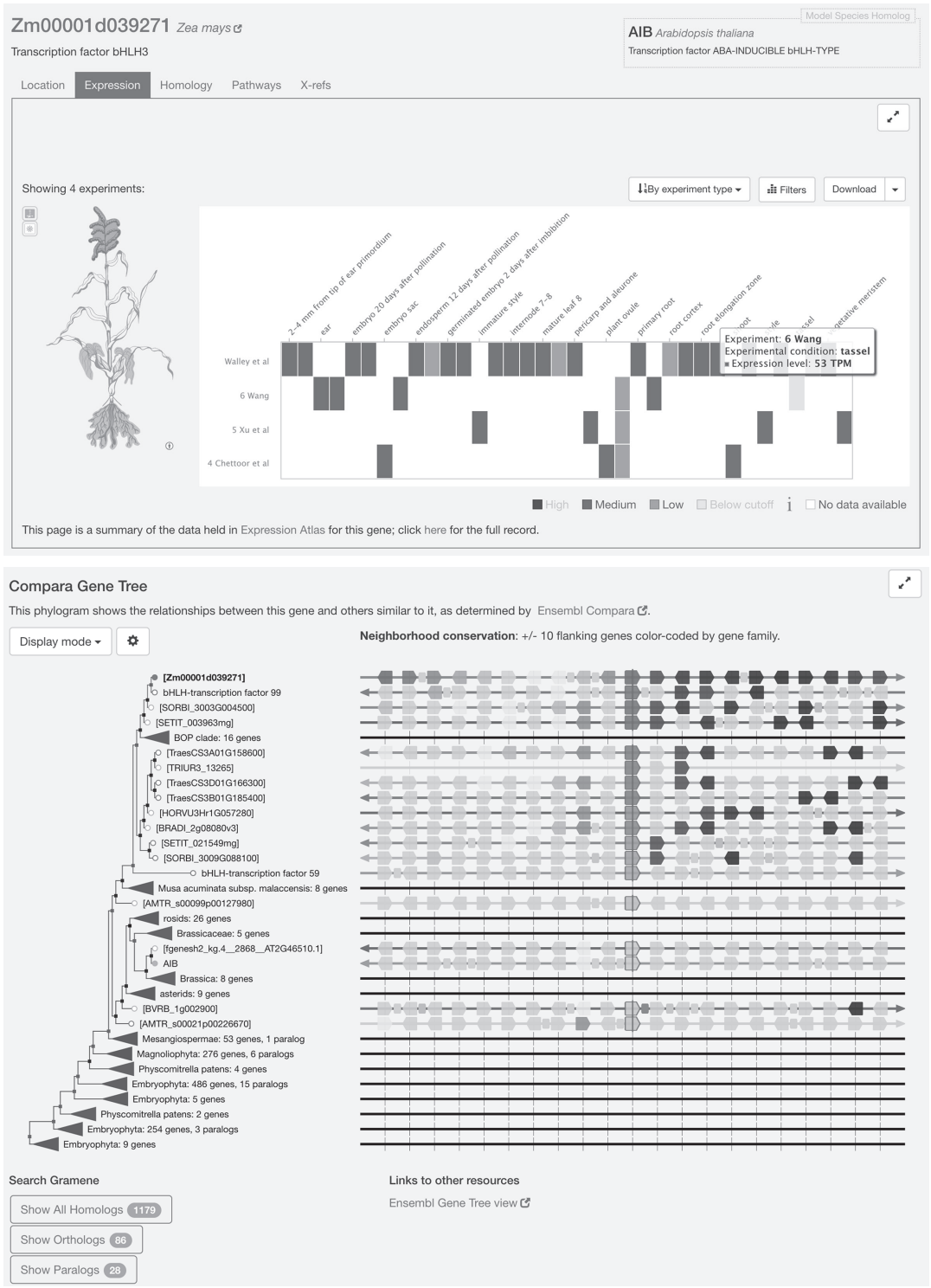


Figure 1. Gramene search views in build 55 (September 2017) are provided as tabs: Location, Expression, Homology, Pathways, and X-refs (or cross-references). (*A*) The Expression view on display shows curated baseline gene expression data from the Expression Atlas project (https://www.ebi.ac.uk/gxa/plant/experiments) for Zm00001d039271, transcription factor bHLH3. (*B*) The neighborhood conservation mode of the Homology view shows a phylogram of ±10 flanking genes color-coded by gene family.

but is limited by sources of variation. Wild relatives of rice, collected from around the world, provide a reservoir of resistance genes that can be transferred to cultivated rice by introgressive breeding. Taking advantage of a 13-species set of I-OMAP consortium reference genomes in OGE Gramene (www.oge.gramene.org) that spans the *Oryzeae* tribe (including wild species in the *Oryza* and *Leersia* genera), we discovered more than 5400 NLR (nucleotide-binding domain and leucine-rich repeat-containing) genes in 28 families (Stein et al. 2018). Rapid diversification of complex haplotypes by gene expansion and loss is typical of NLR genes, contributing to disease adaptation. Applying phylogenetic reconciliation methods to gene trees in these 28 families, we found a 10-fold increase in duplication rates in lineages leading to both Asian and African cultivated rice, consistent with selection for resistance traits before domestication. Most NLR were positionally clustered, often forming complex arrangements of distantly related genes (Fig. 2). Yet, clear orthologous relationships and evidence of conserved underlying haplotype structures could be drawn, even in the most distantly related (~17 MY) species of *Leersia*. In rice and other plants, disease resistance is sometimes conferred by the action of two neighboring NLR genes whose products function as heterodimers. Examining all possible combinations of adjacent NLR pairs, we found that those reminiscent of functionally coupled NLR genes, and therefore belonging to heterogeneous families and configured in a head-to-head (divergently transcribed) arrangement, are significantly more prevalent than expected by chance and are also more likely to be conserved across species than other arrangements. This finding suggests that evolution of haplotype structure may be constrained by underlying regulatory and functional interactions between such putative coupled NLR genes. Furthermore, we found a greater prevalence of putative integrated decoy domains among such pairs, which are thought to function in pathogen recognition by mimicking host targets of pathogenicity factors. Striking variation in domain structure suggests that swapping of various decoy domains contributes to the evolution of haplotype diversity and resistance specificity. This study has opened a treasure trove of potentially novel resistance functions that may help in the future development and sustainability of rice.

PLANT SYSTEMS BIOLOGY

Exploring *Arabidopsis* Gene Regulatory Networks

L. Zhang, A. Olson, Y.K. Lee, B. Wang [in collaboration with S. Brady, UC Davis; M. Frank and B. Shen, DuPont-Pioneer, Inc.]

Plant microRNAs (miRNAs) play an important role in regulating plant development and stress responses by posttranscriptionally repressing the expression of their target genes. To identify upstream regulators of miRNA expression, we generated the *Arabidopsis* miRNA Gene Regulatory Network (ARMIG) using a yeast one-hybrid (Y1H) approach. Using a nearly complete root transcription factor (TF) library, we screened 180 miRNA promoters, their targets, and TFs, which are highly connected within the network, and obtained 5376 protein–DNA interactions (PDIs). ARMIGs are characterized by highly connected genes as "hubs" in the network. The zinc-finger-homeodomain (2F-HD) TFs were identified as a subhub network within the miRNA network. An initial survey of six single loss-of-function mutants identified from existing T-DNA collections revealed no observable root phenotypes, suggesting that the ZF-HD TFs are functionally redundant. To test this hypothesis, we generated multiple loss-of-function mutants using different approaches, including genetic crosses to a single loss-of-function mutant, generation of artificial microRNA (amiRNA), and independent repressor lines (Fig. 3). Combining multiple loss-of-function mutants within a single *Arabidopsis* line, we observed several phenotypes, including altered flower structures and an increased vegetative branching. In addition, we performed transcriptome profiling of ZF-HD TF mutants to characterize differentially expressed genes. The gene-expression profiles and phenotypic properties of the mutants suggested a role for ZF-HD TFs as regulators of developmental transitions. Moreover, the results of this work show that the miRNA gene regulatory network (GRN) can be applied more generally, beyond the root system.

To understand the molecular mechanism of ZF-HD TFs, we use RNA-Seq analysis to reveal gene perturbation between wild-type and loss-of-function mutants, including amiRNA transgenic line, *hb23/31/33/34* mutant, and two independent repressor lines 35s:HB21:SRDX and 35s:HB31:SRDX. In the *hb23/31/33/34* mutant, 91% of differentially

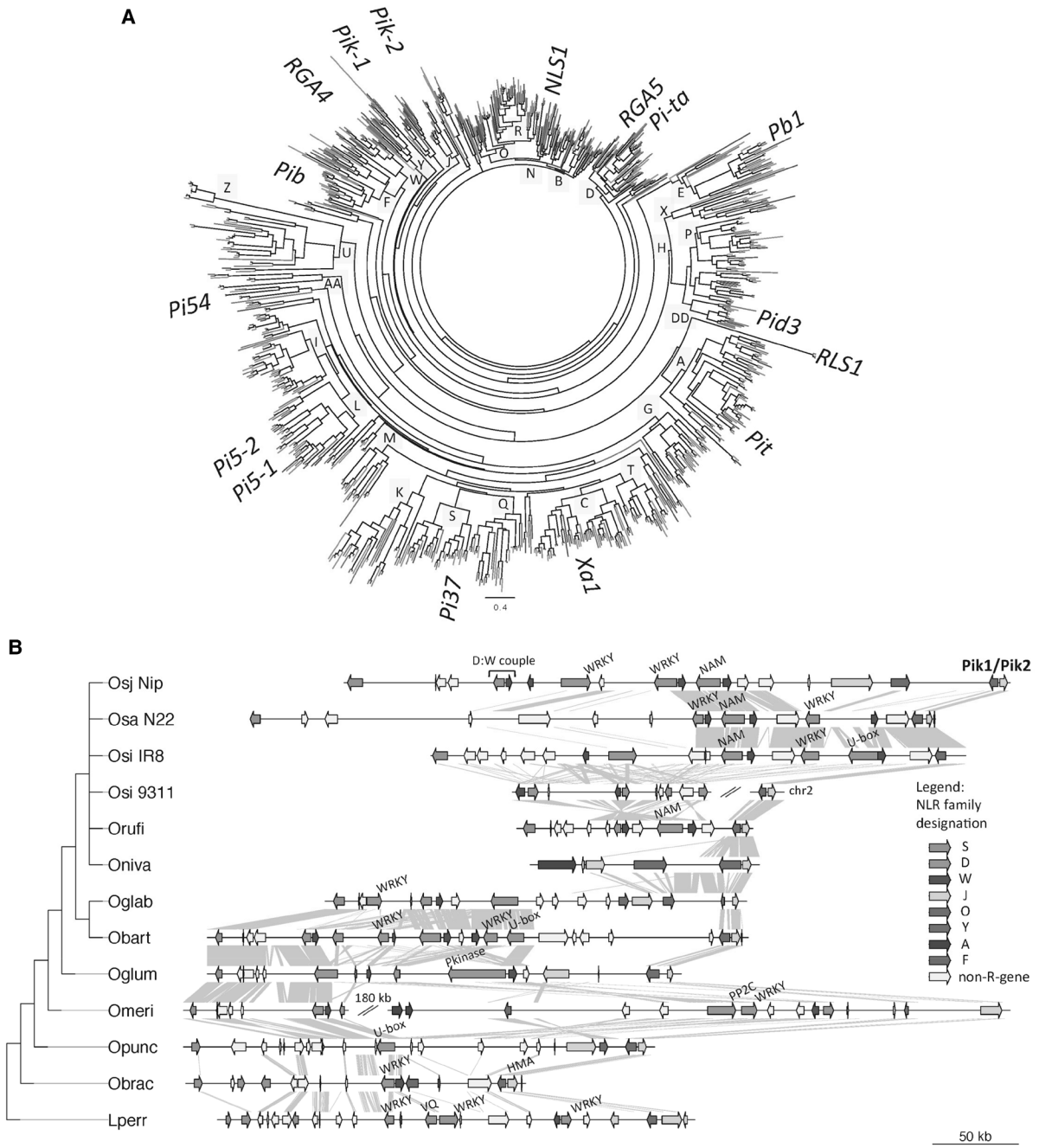


Figure 2. Diversity of NLR genes in the *Oryzeae*. (*A*) Maximum likelihood tree of the NB-ARC domains of 1000 NLR-type disease-resistant genes in three species, *Oryza sativa* vg. indica [93-11], *Oryza punctata*, and *Leersia perrieri*. Designated gene families are lettered. Sixteen characterized NLR genes are indicated, as cited in Online Methods. (*B*) The rice *Pik1*/*Pik2* locus and encompassing heterogeneous NLR cluster showing a pattern of underlying conservation, including a prevalence of putatively coupled gene pairs (W:D pairs in h2h configuration, akin to the rice blast resistance locus, *RGA4*/*RGA5*). Noncanonical carboxy-terminal domains, such as WRKY, which function as integrated decoys that recognize pathogen avirulence gene products, are found in 17 "D" genes as labeled. Region shown corresponds to 11:27,594,830-27,990,888 on the *O. sativa* vg. japonica cv. Nipponbare RefSeq. Species key: Osj Nip, *O. sativa* vg. japonica; Osa N22, *O. sativa* vg. aus [N 22]; Osi IR8, *O. sativa* vg. indica [IR 8]; Osi 9311, *O. sativa* vg. indica [93-11]; Orufi, *O. rufipogon*; Oniva, *O. nivara*; Oglab, *O. glaberrima*; Obart, *O. barthii*; Oglum, *O. glumaepatula*; Omeri, *O. meridionalis*; Opunc, *O. punctata*; Obrac, *O. brachyantha*; Lperr, *L. perrieri*.

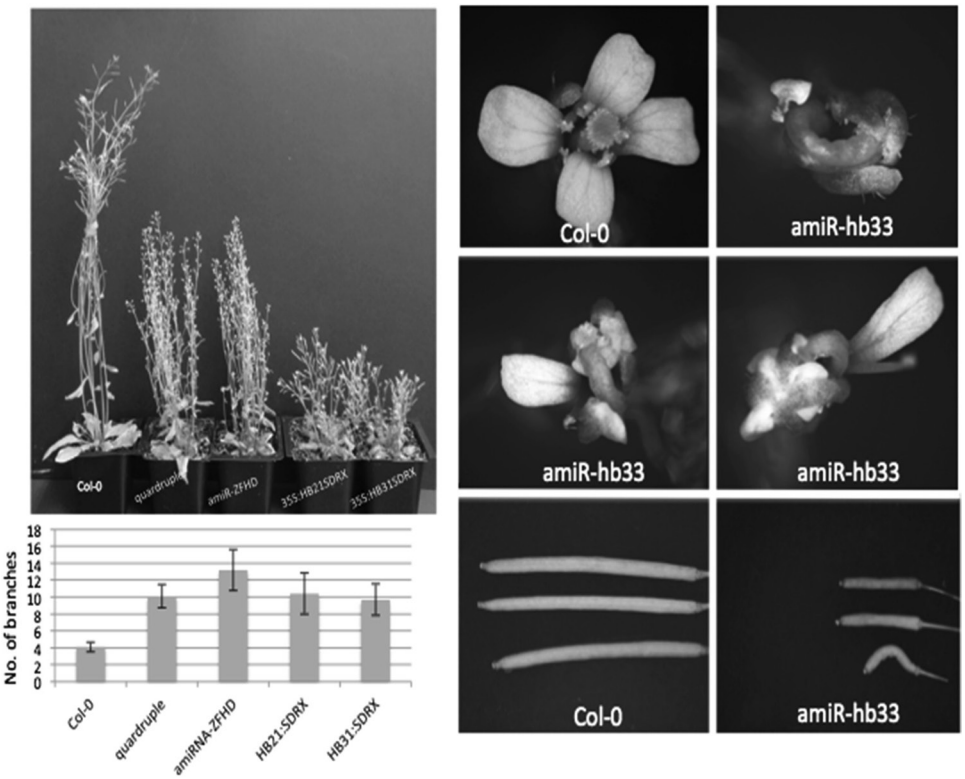


Figure 3. Phenotype of the ZF-HD TF mutant. Branching phenotype of quadruple mutant, amiRNA, and two independent repressor lines compared with wild-type plant Col-0. (*Top left and right*) Lateral branching number of wild-type and mutants.

down-regulated genes are found in the amiRNA line compared with 48% and 53% of differentially expressed genes in the HB31 and HB21 repressor lines, respectively. A similar pattern was observed for differentially expressed putative TF genes. To identify direct targets of ZF-HD TFs at genomic scale, we performed genome-wide chromatin immunoprecipitation (ChIP)-seq analysis of HB34 in young flower tissue. To find out the real target of ZF-HD TFs, we are trying to overlay with ChIP-seq and RNA-Seq data. We found that *HB34*, along with other ZF-HD TFs, negatively regulates the expression of miR157 and positively regulates *SQUAMOSA PROMOTER BINDING–LIKE 10*. In addition, *MIR157D* and *SPL10* are direct targets of HB34, generating a feed-forward loop. This loop constitutes a robust regulatory module that determines plant architecture in *Arabidopsis*. Moreover, phylogenetic analysis and genetic complementation revealed that the ZF-HD TFs play functional roles that are conserved among land plants. To apply this in crop plants, we are screening mutant populations of maize and sorghum.

The basic leucine zipper (bZIP) TFs constitute another highly connected TF family. One member of the family, GBF2, binds many miRNA and miRNA target promoters. As with the ZF-HD TFs, single-gene loss-of-function mutants do not have clear phenotypes. We are in the process of generating multiple loss-of-function mutants by performing genetic crosses between the available single-gene mutant lines.

Dissection of Gene Regulatory Networks Associated with Abiotic Stress Responses

L. Zhang, A. Olson, V. Kumar, S. Kumari, K. Chougule [in collaboration with A-M. Bagman, A. Gaudinier, and S. Brady, UC Davis; M. Frank and B. Shen, DuPont-Pioneer, Inc.]

Nitrogen (N) is an essential micronutrient for plants. Maximizing nitrogen use efficiency (NUE) in plants is critical to the increase in crop production and reduction of negative impacts on the environment. To explore the GRN that controls these processes, we have profiled the transcriptome of maize in response to N limitation,

and, based on this profile, as well as genes known to be involved with nitrogen uptake, assimilation, use, remobilization, and transcriptional regulation in maize, we have used a gene-centered approach to characterize transcription factors that regulated these genes. We are in the process of performing functional genomics screens and computational analyses to model the impact of these genes on plant fitness. This information can be used to inform future breeding strategies.

Exploring a GRN Associated with Response to Cold and Heat

L. Zhang, F. Hu, M. Regulski [in collaboration with S. Anderson and N. Springer, University of Minnesota]

Transposable elements (TEs) were first discovered by Barbara McClintock at Cold Spring Harbor Laboratory (McClintock *Genetics 38:* 579 [1953]). They are also described as "controlling elements," which exist in many eukaryotic genomes and account for 85% of the genome in maize. Recent research indicated TEs are not "junk" DNAs but rather contribute to the regulation of gene expression. Dr. Nathan Springer's group (University of Minnesota) has profiled the genes and TE expression of maize in response to a number of abiotic stress factors (cold, heat, salt, and UV stress). The analysis of TE families inserted within upstream regions of up-regulated genes revealed that TEs were associated with up-regulated gene expression in stress conditions and, in many cases, to the same stress conditions. The analysis of the stress-induced transcripts and proximity of the transposon to the gene suggests that these TEs may provide local enhancer activities that stimulate stress-responsive gene expression (Makarevitch et al. *PLoS Genet 11:* e1004915 [2015]). In collaboration with Dr. Springer's group, we selected 45 candidate genes that are up-regulated in response to either cold or heat, and are exploring the impact of TEs and their effect on response to stress.

Developmental Networks Controlling Inflorescence Architecture in Grasses

Y. Jiao, Y.K. Lee, S. Kumari [in collaboration with Z. Xin and J. Burke, USDA ARS]

The goal of this work is to integrate genetics and genomics data sets to find molecular networks that influence the morphology (architecture) of grass inflorescences (flowers). Because inflorescences bear the fruits and grains that we eat, the genetic and regulatory factors that govern their formation are clearly relevant to important agronomic traits such as grain yield and harvesting ability. In addition to our previous work in maize, we have begun work on sorghum, an important emergent bioenergy crop, which is also used for human consumption in sub-Saharan Africa. The number of grains per panicle is a developmental trait that contributes to sorghum yield. Sorghum flowers comprise one fertile (sessile) and two sterile (pedicellate) spikelets, with only the sessile spikelet producing seed. Using an EMS population, we identified independent multiseeded (*msd*) mutants with both fertile sessile and pedicellate spikelets. A detailed dissection of developmental stages of wild-type and *msd1* revealed that pedicellate spikelets in wild-type do not have floral organs, including ovary, stigma, filament, or anther, whereas the *msd1* mutants generate intact floral organ in both sessile and pedicellate spikelet (Fig. 4). Using a bulk segregant analysis of F2 individuals, we determined that the *msd1* mutations are located within a TCP transcriptional factor. The six causal SNPs found in *msd1* are highly conserved across grass species. The *TCP* gene was found to be differentially expressed during inflorescence development. To characterize the gene networks associated with pedicellate spikelet fertility, we generated whole-genome expression profiling data of floral tissues at four different inflorescence development stages in both wild-type and *msd1*. Preliminary analyses suggest *MSD1* may impact programmed cell death signaling in pedicellate spikelets in wild-type by modulating hormone pathways.

CYBERINFRASTRUCTURE PROJECTS

CyVerse (Formerly The iPlant Collaborative)

L. Wang, Z. Lu, K. Chougule, X. Wang, P. Van Buren, J. Stein, D. Ware [in collaboration with Cold Spring Harbor Laboratory, employing over 100 staff and headquartered at the University of Arizona, principal investigators (PIs) A. Parker, D. Ware, N. Merchant, M. Vaughn, and E. Lyons; dozens of collaborators located at more than 20 institutions]

Our world is changing rapidly. The human population is increasing while arable land and fisheries are decreasing and food cultivation is being diverted for

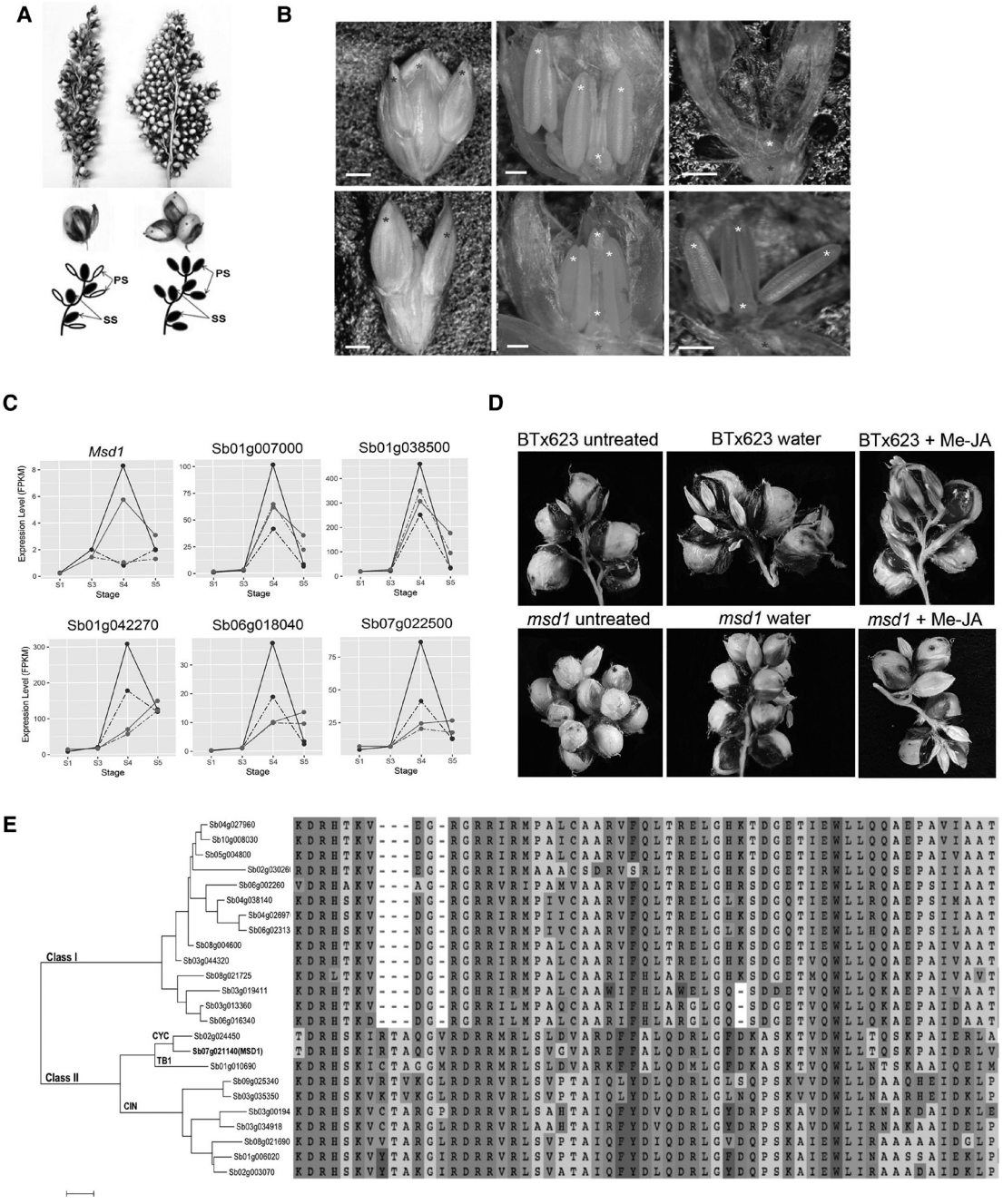


Figure 4. *Msd1* suppresses floral organ PS at developmental stage 4. (*A*) Photographs of primary branches of the panicle (adaxial side) of BTx623 (wild-type, WT) and *msd1*. Images immediately below show terminal spikes: WT has one sessile grain, whereas *msd1* has three grains, one sessile and two pedicellate. The figure provides a schematic representation of a secondary branch depicting SS and PS in WT and *msd1*: *left*, WT; *right*, *msd1*. SS, sessile spikelet; PS: pedicellate spikelet. (*B*) Fully developed (mature) spike of WT (*top* row) and *msd1* (*bottom* row) at stage 5; dissected mature (lemma and palea pulled away) (*middle* column); and mature PS (*right* column). Scale bar, 1 mm. (*C*) Box plot showing jasmonate (JA) concentrations from five young panicles from each of two *msd1* mutants and WT (BTx623). Lines from *bottom* to *top* represent minimum, first quartile, median, third quartile, and maximum; dots indicate outliers. JA levels were lower in the two mutants than in WT. (*D*) Photograph of the adaxial side of a secondary branch of inflorescence during exogenous JA treatment of WT and *msd1-1* plants. The *msd1-1* mutant reverted to the WT phenotype following treatment with methyl jasmonate (Me-JA). (*E*) Polygenetic tree of the 24 genes in TCP transcriptional factor family in sorghum. The *MSD1* gene is bolded, which belongs to the Class II CYC family.

fuel production. Climate instability and energy sustainability are impacting agricultural and ecological systems while concomitant changes in land-use patterns affect global biodiversity. To successfully address these issues, we need to understand how organisms' appearance, physiology, and behavior are shaped by the interactions between their genetic makeup and the environment. Although these global challenges are sobering, the efforts to respond productively will lead to exciting science—provided that the computational infrastructure is in place to handle the necessary data sets, analyses, interpretation of results, and dissemination of knowledge. Advances in biological research technology have enabled scientists to amass unprecedented amounts of data, and many researchers find themselves drowning in this sea of data. Foreseeing this major bottleneck in biological research, the U.S. National Science Foundation (NSF) established the iPlant Collaborative (http://iplantcollaborative.org) in 2008 to develop cyberinfrastructure for life sciences research and democratize access to U.S. supercomputing capabilities. In 2015, iPlant was rebranded as CyVerse to emphasize an expanded mission to serve all life sciences.

Having completed its first nine-year grant, the CyVerse Collaborative has made extensive progress toward meeting these goals. Work in the last year culminated in the development of multiple platforms that are now being used by the research and educational communities. Staff at CSHL contributed directly to some of these cyberinfrastructure platforms or built on existing platforms to provide scientists and educators with ready access to needed software and analysis tools. Within the Ware lab, these platforms include the CyVerse Data Store, Discovery Environment (DE), and Atmosphere. In addition, the Ware lab has successfully built the first federated CyVerse system at CSHL, which supports a dozen high-performance computing (HPC) apps running on a local cluster and storage system. Leveraging the federation system, a workflow management system, SciApps, has been developed for supporting automation of complicated workflows.

The CyVerse Data Store was designed as a distributed storage system for hosting data on the cloud, providing convenient access. The distributed system allows us to "bring the infrastructure to the data" to enable fast computing, and at the same time, reduce the bandwidth required to transfer large amounts of next-generation sequencing (NGS) data. Currently, the CyVerse Data Store consists of two major storage systems with fast internet connections, one at the University of Arizona (UA) and another one at the Texas Advanced Computing Center (TACC). CSHL itself has relatively limited bandwidth relative to TACC and UA, making our institution an ideal case for validation of the CyVerse model of bringing the infrastructure to the data. Because it makes more sense to use local computing resources with CyVerse prebuilt analysis workflows via the CyVerse Agave API than to copy the data to outside clusters (e.g., it takes >10 min to move 1 GB of data from CSHL to the TACC clusters), the Ware lab at CSHL decided to migrate and have successfully migrated the important CyVerse platform components at CSHL. The first target is the Data Store, and a local resource server (WildCat) has been successfully added into the CyVerse storage system, allowing data transfer that is 80 times faster than with outside servers. For the second step, the Ware lab has successfully synchronized a computing cluster with CyVerse's Agave servers at TACC, which allows CyVerse's prebuilt workflows to be run locally on CSHL computing servers. The migration of these important platforms provided a proof-of-concept demonstration of the portability of CyVerse Platforms for enrolling more institutes for efficient data management and possibly use of national computing resources provided with good bandwidth.

The DE is the most visible portal for CyVerse tools and services. This web-based platform supports an "app store" model of user-extensible tools, automated workflows, and data storage. Users can take advantage of existing tools, integrated by CyVerse staff and the user community, or add their own tools to use privately or share. Although users may not be aware of this, the underlying infrastructure provides access to CyVerse's massive Data Store at UA and TACC. Computationally intensive tasks are handled by supercomputers located at TACC and other centers within the Extreme Science and Engineering Discovery Environment (XSEDE). So far, more than 300 tools have been integrated into the DE; these tools enable a broad range of research activities, including genome/transcriptome assembly, annotation, RNA-Seq quantitation, variant detection, genome-wide association studies (GWASs), and phylogenetics. Members of the Ware lab have played important roles in contributing to workflow design, tool integration, validation, science tutorials, and documentation.

Atmosphere is CyVerse's configurable and cloud-enabled computational resource for the plant research community. From Atmosphere's web interface, users can launch a virtual machine (VM) with preconfigured working environments and precustomized, ready-to-use software. Users can also create their own applications and environments as VMs and share them with others via Atmosphere. As with the DE, Atmosphere is a gateway to CyVerse's core infrastructure resources, including the high-performance grid computing environment and big-data storage system.

We have also developed SciApps, a web-based platform, for reproducible bioinformatics workflows. For creating a workflow, each analysis job is submitted, recorded, and accessible through the web portal. Part or all of the series of recorded jobs can be saved as reproducible, sharable workflows for future execution with the original or modified inputs and parameters. The platform is designed to automate the execution of modular Agave apps and support executing workflows on either local clusters or the cloud.

The success of genome research depends on our ability to accurately assemble, annotate, and derive meaning from sequence data; however, the extremes of genome size, polyploidy, diversity, and repeat content push the limits of the algorithms, expertise, and computational power currently available to researchers. In response, CyVerse is fostering a community effort to identify best practices and state-of-the-art tools, install them, optimize their performance on the nation's most powerful supercomputers, and make them available as free online resources. Over the last three years, the CyVerse DE has matured to provide a comprehensive set of tools and services for sequence handling, read alignments, RNA-Seq profiling, and de novo genome and transcriptome assembly. To extend these capabilities, we have incorporated MAKER-P, a standardized, portable, and easy-to-use plant-genome annotation engine with built-in methods for quality control. As part of this effort, MAKER-P was specifically optimized to take advantage of the parallel computing environment of the TACC Lonestar cluster, and is now a supported module. Performance testing showed that MAKER-P provides high-quality, full-fledged annotation pipelines on even the largest plant genomes in a matter of hours. MAKER-P is currently available for use as an Atmosphere and Jetstream image.

A major mission of CyVerse is to promote adoption of the cyberinfrastructure through training workshops and outreach at academic institutions and scientific meetings. In 2017, members of the Ware lab participated as instructors in several "Big Data" workshops focused on transcriptomics and annotation using the DE, Atmosphere, and SciApps platforms.

KBase: Department of Energy Systems Biology Knowledgebase

V. Kumar, S. Kumari, J. Thomason [in collaboration with DOE National laboratories and led by PI A. Arkin, Lawrence Berkeley National Laboratory (LBNL); co-PIs C. Henry, Argonne National Laboratory (ANL) and R. Cottingham of Oak Ridge National Laboratory (ORNL). As Plants Science Lead for KBase, D. Ware continues to informally serve as a co-PI on the project]

The Systems Biology Knowledgebase (Kbase; www.kbase.us) is an open-access, collaborative platform for systems biology of plants, microbes, and their communities, and has two primary goals. The scientific goal is to produce predictive models, reference data sets, and analytical tools, and show their utility in U.S. DOE biological research relating to bioenergy, carbon cycle, and the study of subsurface microbial communities. The operational goal is to create the integrated software and hardware infrastructure needed to support the creation, maintenance, and use of predictive models and methods for the study of microbes, microbial communities, and plants.

KBase was conceived from the beginning as a knowledgebase that would bring together relevant computational systems biology tools and data for microbes, plants, and interactions between the two. To support this, KBase's data model integrates reference data and shared user data for both microbes and plants so that results for one organism can be applied to others, helping to accelerate the pace of systems biology research. KBase also adheres to the FAIR (findable, accessible, interoperable, reusable) data principles endorsed by many funding agencies and scientific organizations.

KBase is designed to support large-scale analysis by providing ways to upload large individual data files and large collections of data, as well as enabling bulk processing. During 2017, the data upload utility was significantly revised to improve usability, performance, and support for additional data types. KBase now supports upload and import of sequencing reads

(paired end, single end, SRA), assemblies, genomes (GenBank files), metabolic models and media, and phenotypes through its Narrative "staging" interface. Users can upload data sets to their KBase account from their computer or an online site with a publicly available URL (http, ftp, Dropbox, or Google Drive). New capabilities in KBase for bulk upload and execution make it possible to upload large data sets and quickly run them through sophisticated analysis workflows. The members of the Ware lab contributed directly to the product design and development of user interface and validation of the growing capability. They also actively support the issue resolution and community as they use these tools.

KBase has a range of analysis tools and data resources that enable the plant science community to gain insight into the evolution of genes and genomes, profile transcriptomes, perform genome functional modeling with metabolic networks, and identify differential expression between tissues, developmental stages, environmental conditions, and genetic backgrounds (Fig. 5). These capabilities are directly relevant to important DOE research targets such as optimizing biomass production in biofuel feedstocks.

The underlying apps enable users to analyze large genomes and NGS data sets. NGS reads can be trimmed and quality-checked and then assembled and run through RNA-Seq pipelines and downstream analysis tools to study differential expression. Genomes can be annotated and then used as input to KBase's metabolic modeling tools, which can reconstruct models for eukaryotes as well as prokaryotes. Another analysis workflow starts with sequence alignment and proceeds to phylogenetic tree reconstruction to elucidate the evolutionary relationships among multiple plant species.

To support this work, KBase has integrated several key sources of eukaryotic reference data, such as annotated plant genomes from JGI's gold-standard resource, Phytozome, to give KBase users immediate access to this valuable community resource and enable them to integrate it with their own data. Integration of these resources into KBase allows users to not only apply KBase analyses to their organism of specific interest but also compare the results among different organisms and combine them. KBase also supports user upload of large NGS reads and genome sequences that can be analyzed together with the public data. Together, these tools—some of which are not available elsewhere, such as the plant-microbiome metabolic modeling capability—and data resources bring immediately useful sophisticated analysis capabilities

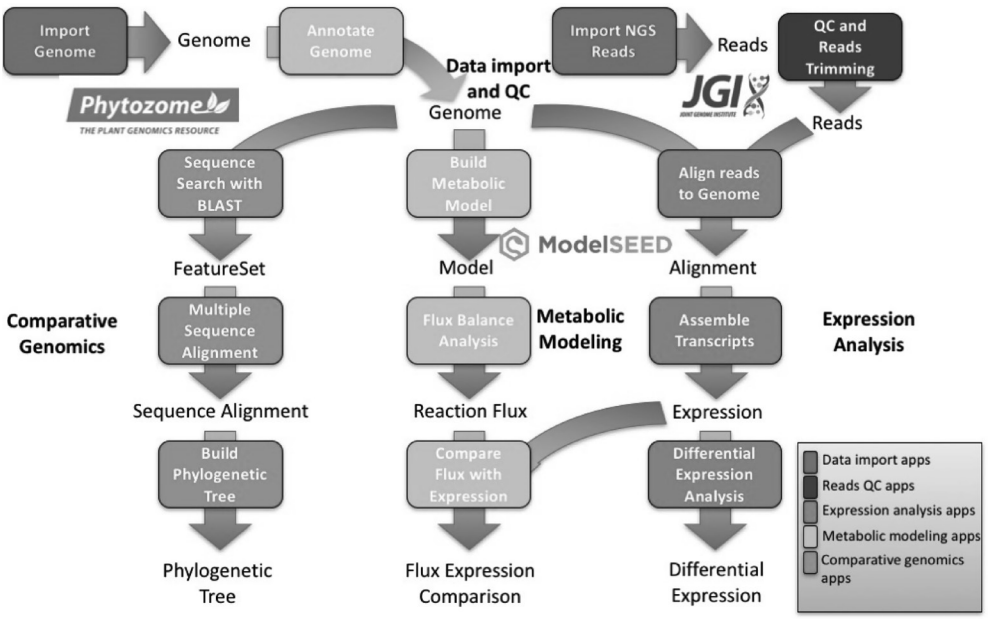


Figure 5. The data import and analysis apps (in colored boxes) that can be used to build plant analysis workflows in KBase. NGS, next-generation sequencing; QC, quality control.

into the hands of KBase users. Combined with the platform infrastructure, these integrated resources significantly lower the bar for future community development and integration of plant systems biology tools and data.

Another active area of development is the expression analysis toolkit, which includes support for original and new Tuxedo tools and is being expanded and improved. STAR (Spliced Transcripts Aligner to a Reference) is an ultrafast RNA-Seq aligner tool to align reads to genomes; it can map reads of any length and detect canonical and noncanonical splice junctions and fusion transcripts. DESeq2 is a counts-based differential expression tool that generates an expression matrix based on the total counts of a gene, including all of its isoforms. QualiMap provides a comprehensive multi-BAM (binary alignment map) QC report.

To offer additional biological insight into differentially expressed genes, KBase has developed additional tools (some in beta) that will support downstream analyses such as functional enrichment. These tools characterize the molecular functions of differentially expressed genes based on Gene Ontology and SEED terms by comparing a list of differentially expressed genes against the rest of the genome to identify overrepresented functions and apply statistical methods to test for enrichment of each annotated gene set. Several apps for downstream analysis of expression data are already available in a preliminary form, with improvements under way. The expression data generated by Cuffdiff, Ballgown, and DESeq2 can be used by clustering apps such as Hierarchical, K-Means, and WGCNA to enable users to analyze patterns of gene expression by grouping expression data. The clusters generated by these apps can be viewed as an interactive heat map. The expression matrices generated by the RNA-Seq apps can be fed to the metabolic modeling apps Compare Flux with Expression and View FBA Expression Comparison to compare reaction flux with gene expression and identify the pathways in which expression and predicted flux agree or conflict.

KBase is actively engaging the external community to help us improve our tools and workflows for plant science, including support for large-scale reads upload and analysis, plant genome annotation, functional genomic clustering and enrichment, physiological modeling and variation, and trait-based modeling analysis. The Ware lab members have actively engaged the community through various channels such as presentations, talks, posters, and demonstrations during major community events such as Plant Biology, Plant and Animal Genomes, Genome Informatics, Biology of Genomes. In addition, they participated in metabolic modeling and transcriptomics workshops at the University of Chicago and Cold Spring Harbor Laboratory. We welcome the community's feedback on our current tools and your suggestions on what new functionality we should add, and invite you to share your plant science Narratives with the community.

PUBLICATIONS

Friesner J, Assmann SM, Bastow R, Bailey-Serres J, Beynon J, Brendel V, Buell CR, Bucksch A, Busch W, Demura T, et al. 2017. The next generation of training for *Arabidopsis* researchers: bioinformatics and quantitative biology. *Plant Physiol* **175:** 1499–1509.

Jiao Y, Peluso P, Shi J, Liang T, Stitzer MC, Wang B, Campbell MS, Stein C, Wei X, Chin CS, et al. 2017. Improved maize reference genome with single-molecule technologies. *Nature*. **546:** 524–527.

Majoros WH, Campbell MS, Holt C, DeNardo EK, Ware D, Allen AS, Yandell M, Reddy TE, et al. 2017. High-throughput interpretation of gene structure changes in human and nonhuman resequencing data, using ACE. *Bioinformatics* **33:** 1437–1446.

Majsec K, Bhuiyan NH, Sun Q, Kumari S, Kumar V, Ware D, van Wijk KJ, et al. 2017. The plastid and mitochondrial peptidase Network in *Arabidopsis thaliana*: a foundation for testing genetic interactions and functions in organellar proteostasis. *Plant Cell* **29:** 2687–2710.

Turco GM, Kajala K, Kunde-Ramamoorthy G, Ngan CY, Olson A, Deshphande S, Tolkunov D, Waring B, Stelpflug S, Klein P, et al. 2017. DNA methylation and gene expression regulation associated with vascularization in *Sorghum bicolor*. *New Phytol* **214:** 1213–1229.

In Press

Stein JC, Yu Y, Copetti D, Zwickl DJ, Zhang L, Zhang C, Chougule K, Gao D, Iwata A, Goicoachea JL, et al. 2018. Genomes of 13 domesticated and wild rice relatives highlight genetic conservation, turnover and innovation across the genus *Oryza*. *Nat Genet* **450:** 285–296.

QUANTITATIVE BIOLOGY

Gurinder S. "Mickey" Atwal and colleagues are applying insights from the physical and computational sciences to the study of population genetics and human disease. The Atwal lab has modeled the process by which genetic variants, or alleles, have evolved in the last 100,000 years of human history. This has recently led to surprising insights about the role of *p53*, a master tumor suppressor gene, in female fertility and furthered our understanding of how complex gene networks evolve. The lab has analyzed the comparative genomics and physical organization of cancer-related genes and their role in mediating tumorigenesis across numerous tissue types. Recently, they have begun to focus efforts on understanding cancer genome evolution on shorter time scales by analyzing nucleotide sequences from single cells.

Ivan Iossifov focuses on the development of new methods and tools for genomic sequence analysis and building and using molecular networks, and he applies them to specific biomedical problems. He studies the genetics of common diseases in humans using two main tools: next-generation sequencing and molecular networks representing functional relationships among genetic loci. These approaches in combination enable the kind of large-scale studies necessary for furthering our understanding of the complex etiology of disorders such as autism, bipolar disorder, and cancer.

Justin Kinney completed his Ph.D. in Physics at Princeton University in 2008 and began his term as a Quantitative Biology Fellow in 2010. His research focuses on developing next-generation DNA sequencing as a tool for dissecting the structure and function of large macromolecular complexes. Of particular interest to his lab is the biophysical basis of transcriptional regulation—how simple interactions between proteins and DNA allow promoters and enhancers to modulate genes in response to physiological signals.

In 2010, Kinney and colleagues published a paper demonstrating Sort-Seq, a novel sequencing-based method that can measure the functional activity of hundreds of thousands of slightly mutated versions of a specific DNA sequence of interest. Using a novel information-theoretic analysis of the resulting data, Kinney et al. were able to quantitatively measure, in living cells, the protein–DNA and protein–protein interactions controlling mRNA transcription at a chosen promoter. Kinney continues to develop this approach using a combination of theory, computation, and experiment. From a biological standpoint, Sort-Seq allows researchers to investigate important but previously inaccessible biological systems. Kinney's lab is currently using Sort-Seq to address open problems in transcriptional regulation, DNA replication, and immunology. These experiments also present new challenges for the field of machine learning, and a substantial fraction of Kinney's efforts are devoted to addressing the theoretical and computational problems relevant to the analysis of Sort-Seq data.

Alexander Krasnitz and colleagues use mathematical and statistical tools to discover key genetic elements involved in cancer and to understand how cancer cells evolve. Array-based comparative genome hybridization, a technique honed in the Wigler lab, and, more recently, sequencing experiments, have revealed subtle patterns of frequent and widespread aberration in cancer genomes. Krasnitz hypothesizes that recurrent, aberrant genomic loci observed in a range of cancer types are under selection and therefore are enriched in important cancer genes. He has developed a novel, comprehensive methodology to discover such "cores" and has used it to analyze multiple genome data sets in breast, liver, ovarian, and prostate cancer. The results have been shared with cancer biology labs across CSHL, and they have been a key enabling agent of functional studies using mouse models and RNA interference. Krasnitz has begun to apply these novel statistical tools to

the latest generation of experimental data, which have characterized tumor samples down to the level of single cells. By interpreting single-cell genomes, he and his colleagues seek to learn how specific tumors evolve and how cancer cells migrate to invade adjacent tissues and metastasize.

There is increasing evidence that rare and unique mutations have a significant role in the etiology of many diseases such as autism, congenital heart disease, and cancer. **Dan Levy**'s group develops algorithms to identify these mutations from large, high-throughput data sets comprising thousands of nuclear families. After earlier working with high-resolution comparative genomic hybridization (CGH) arrays, Levy's group now uses targeted sequence data. Levy has developed methods for identifying de novo mutations (i.e., those seen in a child but not in his or her parents) by simultaneously genotyping the entire family; the team is currently focused on building algorithms to detect copy number variants and multiscale genomic rearrangements. Although their copy number methods are based on "read" density, there are classes of mutations that require analysis at the level of the read. Thus, they are developing algorithms to identify insertions, deletions, inversions, transpositions, and other complex events. Other projects in the Levy lab include analysis of single-cell RNA, phylogenetic reconstruction from sparse data sets, and disentangling haplotypes from sperm and subgenomic sequence data.

The **David McCandlish** lab develops computational and mathematical tools to analyze and exploit data from high-throughput functional assays. The current focus of the lab is on analyzing data from so-called "deep mutational scanning" experiments. These experiments simultaneously determine, for a single protein, the functional effects of thousands of mutations. By aggregating information across the proteins assayed using this technique, the team seeks to develop data-driven insights into basic protein biology, improved models of molecular evolution, and more accurate methods for predicting the functional effects of mutations in human genome sequences.

Critically, these data also show that the functional effects of mutations often depend on which other mutations are present in the sequence. The team is developing new techniques in statistics and machine learning to infer and interpret the complex patterns of genetic interaction observed in these experiments. Their ultimate goal is to be able to model these sequence–function relationships with sufficient accuracy to guide the construction of a new generation of designed enzymes and drugs and to be able to predict the evolution of antigenic and drug resistance phenotypes in rapidly evolving microbial pathogens.

Michael Schatz is a computational biologist and an expert at large-scale computational examination of DNA sequencing data, including the alignment, assembly, and analysis of next-generation sequencing reads. These methods have been used to reconstruct the genomes of previously unsequenced organisms, probe sequence variations, and explore a host of biological features across the tree of life. Recent improvements in sequencing technologies are challenging our capacity to store and analyze the huge volume of DNA sequence data being generated. Consequently, Schatz is particularly interested in capitalizing on the latest advances in distributed and parallel computing, especially cloud computing technologies, to advance the state of the art in bioinformatics and genomics. In a recent breakthrough, Schatz was able to create a hybrid software-based solution to eliminate errors in so-called third-generation sequencing. This makes it remarkably easier to compile, align, and analyze full genome sequences.

Modern genomic technologies make it relatively easy to generate rich data sets describing genome sequences, RNA expression, chromatin states, and many other aspects of the storage, transmission, and expression of genetic information. For many problems in genetics today, the limiting step is no longer in data generation, but in integrating, interpreting, and understanding the available data. Addressing these challenges requires expertise both in the practical arts of data analysis and the theoretical underpinnings of statistics, computer science, genetics, and evolutionary biology.

Adam Siepel and colleagues addresses various problems in genomics and evolution using tools and techniques from computer science and statistics. Over the years, Siepel's research has touched on diverse topics, including the identification of recombinant strains of HIV, the discovery of unannotated human genes, the characterization of conserved regulatory elements in mammalian genomes, and the estimation of the times in early human history when major population groups first diverged. General themes in the group's work are the development of precise mathematical models for the complex processes by which genomes evolve over time and the use of these models to both peer into the past and address questions of practical importance for human health. Recently, they have increasingly concentrated on research at the interface of population genomics and phylogenetics, with a particular focus on humans and the great apes. They also have an active research program in computational modeling and analysis of transcriptional regulation in mammals and *Drosophila*, in close collaboration with Charles Danko and John Lis at Cornell University.

These investigators are all part of the **Simons Center for Quantitative Biology** (SCQB), CSHL's home for mathematical, computational, and theoretical research in biology. The central idea behind the SCQB is to place researchers trained in mathematics, physics, computer science, and other quantitative fields on the frontlines in biology, working shoulder to shoulder with experimentalists. In addition to collaborative work, SCQB researchers pursue independent research in algorithms, machine learning, statistical genetics, molecular evolution, and other areas. The ultimate goal of the Center is to promote the use of quantitative methods to enable groundbreaking research across a wide variety of biological domains, including human genetics, cancer, plant biology, and neuroscience. The SCQB was created in 2008 with a major donation from the Simons Foundation, as well as gifts from the Starr Foundation and Lavinia and Landon Clay. Adam Siepel was recruited to serve as Chair of the SCQB in 2014.

QUANTITATIVE BIOLOGY

G.S. Atwal J. Carter L. Kolla
P. Gilbo R. Utama
K. Grigaityte

Fueled by data generated from recent technological developments in DNA sequencing, the Atwal lab primarily develops mathematical models and computational methods to tackle problems in cancer genomics and immunology. Recently, we have focused efforts on understanding the cellular heterogeneity and evolution of the tumor microenvironment, and how we may leverage this knowledge to inform effective cancer immunotherapeutics in breast cancers.

This year, we have continued to press ahead with a number of cancer-related projects and made significant progress on nascent single-cell studies of the human adaptive immune system. We have been following up on the investigation of ectopic expression of germline genes, transcribing genetic loci exclusively expressed in testes/ovaries in various tumors. Previous work in our lab has established the pleiotropic and ancient role of the TP53 pathway in germline development. Preliminary results have also identified ectopic expression of germline *piwi* genes in samples of glioblastoma multiforme extracted from the Cancer Genome Atlas. However, the landscape of ectopically expressed germline genes is unknown, and their functional impact on cancer development remains elusive. We have continued an integrative study that sought to determine the ectopic expression of germline transcripts in glioblastoma multiforme and breast cancer, elucidating the predictive value of germline molecular programs as diagnostic markers and immunogenic therapeutic targets. Taha Merghoub and Jedd Wolchok at Memorial Sloan Kettering Cancer Center (MSKCC) continue to collaborate with us on this project. We initiated a new collaboration with Betsy Barnes at the Feinstein Institute to test candidate peptides as vaccines in mouse models of triple-negative breast cancer.

A major and recent research focus of our lab is the investigation of cancer evolution using sequencing data derived from single-cell genomics. Although the timescales vary by many orders of magnitude, the mathematical tools of population genetics, originally devised to model molecular evolution over millions of years, can be repurposed to understand single-cell evolution of tumors in the lifetime of an individual. In particular, we have been addressing the ubiquitous problem of how many cells and how much read coverage we need before the inferred cell phylogeny accurately reflects the evolutionary history of the tumor. In addition, our lab has continued research in the use of information theory and other machine-learning tools in addressing the tsunami of data generated by next-generation sequencing.

Jason Carter, an M.D./Ph.D. candidate at Stony Brook University, joined the lab in the fall after rotating in the lab the previous year to continue computational research in the adaptive immune repertoire. Phil Gilbo, a resident physician at Northwell, joined the lab as a visiting clinical fellow to develop predictive RNA-based models of cancer patients undergoing immunotherapy. Likhitha Kolla, an URP student, worked on the research mentorship of postdoctoral scientist Raditya Utama. Mickey Atwal won the 2017 Winship Herr Prize for Excellence in Teaching in the Watson School.

Single-Cell Sequencing of the Adaptive Immune System

K. Grigaityte, J. Carter [in collaboration with Juno Therapeutics]

A diverse T-cell repertoire is a critical component of the adaptive immune system, providing protection against invading pathogens and neoplastic changes, relying on the recognition of foreign antigens and neoantigen peptides by T-cell receptors (TCRs). However, the statistical properties of the αβ T-cell repertoire in healthy individuals has remained poorly characterized, in large part because of the laborious task of sequencing single T cells in a high-throughput fashion.

Over the last year, we analyzed a high-throughput data set of full-length, high-quality, paired αβ sequences ($n = 205{,}950$) from peripheral blood samples of five healthy individuals (three males, two females, ages 33–69) acquired through a recently developed microfluidic method of single-cell RNA sequencing in emulsion

droplets. The sequenced T cells were further stratified into $CD4^+$ ($n = 73{,}495$) and $CD8^+$ ($n = 30{,}321$) subtypes, based on paired sequence tags introduced by labeling with DNA-conjugated antibodies. A critical finding from our analysis was that the paired αβ TCR repertoire cannot necessarily be directly inferred by observing the repertoire of one chain alone. That is, although single-chain bulk sequencing provides valuable information regarding the TCR repertoire, it is not equivalent to the paired repertoire and does not fully capture TCR diversity. This finding has important clinical relevance because almost all previous studies on T-cell clonal changes in cancer immunotherapy have made their conclusions using the β-chain sequencing alone. In addition, we showed that the T-cell clone size distribution of every subject shows a power-law behavior with heavier tails of $CD8^+$ T-cell distribution compared with $CD4^+$, likely the result of higher proliferative activity of $CD8^+$ compared with $CD4^+$ T cells. We observed 26 shared paired αβ CDR3 sequences between any two individuals. Given the estimated theoretical diversity of 10^{15} unique TCRs, and the small sample size that could be obtained experimentally compared with the full population of 10^{11} T cells in the human body, we would expect to observe no TCR sharing among individuals. Considering that the probability of TCR sharing between individuals in our data set is 10^{-7}, the existence of 26 shared paired αβ CDR3 suggests that (i) the total diversity of TCRs is less than 10^{15}, and (ii) V(D)J recombination is not a random process and unknown mechanisms must influence gene selection, as well as insertion and deletion content during T-cell maturation in the thymus.

To conclude, our results stress the importance of sequencing αβ pairs to accurately describe and understand T-cell diversity by emphasizing the discrepancies between single and paired αβ chain repertoires. Not only do these findings have important implications for understanding the basic biology that underlies the generation and maintenance of the TCR repertoire, but they also could be particularly critical for developing successful immunotherapeutic approaches and correctly assessing patient response to treatment.

Genomics of the Tumor Microenvironment in Breast Cancer

R. Utama [in collaboration with City of Hope]

The breast tumor microenvironment of primary and metastatic sites is a complex milieu of differing cell populations. However, the genomic expression landscape of the tumor stroma and its role in mediating cancer progression and informing effective therapies is not well understood. Here, we obtained and cultured 52 cell-sorted stromal tissue samples performed by the Lee group at City of Hope across 37 patients from normal, primary tumor, and metastatic tumor sites. Deep RNA-sequencing (RNA-Seq) was performed on $poly(A)^+$ transcripts at the CSHL sequencing facility using a multiplexed barcoding sequencing strategy to increase yield and obviate batch effects. A conservative linear model of expression covariates revealed significant ($q < 0.05$) differential expression of protein-coding genes and long noncoding RNAs (lncRNAs) in the stroma (80 transcripts in normal vs. primary and 108 transcripts in primary vs. metastatic). The majority of these genes was found to be associated with pattern formation, embryogenesis, and the epithelial–mesenchymal transition. Among the genes that were silenced in the epithelial cells, we initiated some immunohistochemistry (IHC) staining for C2orf88, ALOX5AP, and UNC5C at CSHL histology core, which are important in protein kinase A (PKA) binding, immune response, apoptosis, and neural development. We plan to supplement these results with comprehensive single-cell study of stromal cell heterogeneity, capturing the distribution of exclusive stromal genes inside breast tumor. Furthermore, survival analyses of 827 luminal breast tumor samples from TCGA showed the predictive power of stromal gene expression in informing clinical outcome. Together, these results highlight the evolution of stromal gene expression in breast tumors and suggest novel therapeutic strategies targeting the microenvironment.

PUBLICATIONS

Grigaityte K, Carter JA, Goldfless SJ, Jeffery EW, Hause RJ, Jiang Y, Koppstein D, Briggs AW, Church GM, Vigneault F, et al. 2017. Single-cell sequencing reveals αβ chain pairing shapes the T cell repertoire. bioRxiv doi: 10.1101/213462.

In Press

Alexander J, Kendall J, McIndoo J, Rodgers L, Aboukhalil R, Levy D, Stepansky A, Sun G, Chobardjiev L, Riggs M, et al. 2018. Utility of single-cell genomics in diagnostic evaluation of prostate cancer. *Cancer Res* **78:** 348–358.

GENETIC VARIANTS LINKED TO AUTISM TRAITS

I. Iossifov Y-h. Lee S. Marks A. Munoz B. Yamrom

The bulk of our work in 2017 was in analyzing the large data set of whole-genome sequencing data generated from approximately 2400 of the Simons Simplex Collection (SSC) families. These data are a rich resource that we use in numerous projects. In addition, we continue our analysis of the existing whole-exome and single-nucleotide polymorphism (SNP) genotyping data set available for the SSC and AGRE collection (a set of approximately 1000 multiplex families). Below, the abstracts of four projects that are in submission or near-submission status are listed and our current efforts in studying the role of de novo noncoding variants, rare structural rearrangements, and common variants in autism's etiology are shown.

De Novo Indels within Introns Contribute to ASD Incidence

A. Munoz, B. Yamrom, Y-h. Lee, S. Marks, I. Iossifov
[in collaboration with P. Andrews, K-T. Lin, Z. Wang, A. Krainer, M. Wigler, CSHL; B. Darnell, HHMI, Rockefeller University]

Whole-exome sequencing (WES) has allowed us, in the last 10 years, to make remarkable progress in our understanding of the genetics of autism. But there must be major genetic contributions to the disorder that are unexplained. Through whole-genome sequencing, additional types of genetic variants like noncoding variants, small copy number variants, and structural rearrangements might be observed. Because they are so abundant, analysis of which ones are functional is important but challenging. Here, we analyzed whole-genome sequencing data from 510 of the SSC's quad families and focused our attention on intronic variants. In the introns of 546 high-quality autism target genes, we identified 63 de novo indels in the 510 affected children, whereas we found only 37 such events in their unaffected siblings. The difference of 26 events is significantly larger than expected (p-value = 0.01) and using reasonable extrapolation shows that de novo intronic events can contribute to as much as 20% of simplex autism. The significance increases if we restrict to the half of the autism targets that are intolerant to damaging variants in the normal human population and, hence, expected to be even more enriched for autism genes. For these 273 targets, the difference shrinks from 26 to 23 (p-value of 0.005). There was no significant difference in the number of de novo intronic substitutions in these genes or in de novo intronic indels or substitutions in any of the control sets of genes analyzed. The excess of de novo intronic events was seen in introns that separate two coding exons and not in introns separating 5′ UTR exons. A draft manuscript describing these observations in more detail has been submitted to bioRxiv. Since the draft was submitted, we were able to confirm these findings in a larger set of 1886 families from SSC: For the 273 protected autism target genes, we observed 186 and 127 in affected and unaffected children, respectively (p-value of 0.0003).

Recurrent Variant Transmissions Contribute to ASD

I. Iossifov [in collaboration with M. Wigler, CSHL; A. Buja, A. Krieger, University of Pennsylvania; K. Ye, Albert Einstein College of Medicine]

We develop a method of analysis (A2DS) that provides statistical evidence that transmission of shared variants contributes to a disorder. Using a standard measure of genetic relation, test individuals are compared with a cohort of discordant sib-pairs (CDS) to derive a comparative similarity score. We ask whether the test individuals are more similar to the affected than the unaffected siblings. Statistical significance is judged by randomly permuting the affected status in the CDS. In the analysis of published genotype data from the SSC and the AGRE cohorts of children affected with autism spectrum disorder (ASD), we find strong statistical evidence that the affected are more similar to each other than to the unaffected (p-value < 0.0001). On the other hand, unaffected siblings, or parents, are not more similar to the unaffected

siblings than they are to the affected sibling. These results do not depend on ethnic matching or gender.

Damaging De Novo Mutations Diminish Motor Skills in Children on the Autism Spectrum

I. Iossifov [in collaboration with M. Wigler, CSHL; A. Buja and A. Krieger, University of Pennsylvania; K. Ye, Albert Einstein College of Medicine; N. Volkovsky, A.E. Lash, Simons Foundation; C. Lord, Weill Cornell Medical College]

In individuals with ASD, de novo mutations have previously been shown to be significantly correlated with lower IQ, but not with the core characteristics of ASD—deficits in social communication and interaction, and restricted interests and repetitive patterns of behavior. We extend these findings by demonstrating in the SSC that damaging de novo mutations in ASD individuals are also significantly and convincingly correlated with measures of impaired motor skills. This correlation is not explained by a correlation between IQ and motor skills. We find that IQ and motor skills are distinctly associated with damaging mutations, and, in particular, motor skills are a more sensitive indicator of mutational severity than is IQ, as judged by mutational type and target gene. We use this finding to propose a combined classification of phenotypic severity: mild (little impairment of either), moderate (impairment mainly to motor skills), and severe (impairment of both IQ and motor skills).

MUMdex: MUM-Based Structural Variation Detection

I. Iossifov, S. Marks [in collaboration with P.A. Andrews, J. Kendall, Z. Wang, D. Levy, M. Wigler, CSHL; L. Muthuswamy, New York Genome Center]

Standard genome sequence alignment tools, primarily designed to find one alignment per read, have difficulty detecting inversion, translocation, and large insertion and deletion events. Moreover, dedicated split read alignment methods that depend only on the reference genome may misidentify or find too many potential split read alignments because of flaws in the reference genome.

We introduce MUMdex, a maximal unique match (MUM)-based genomic analysis software package consisting of a sequence aligner to the reference genome, a storage-indexing format, and analysis software. Discordant reference alignments of MUMs are especially suitable for identifying inversion, translocation, and large indel differences in unique regions. Extracted population databases are used as filters for flaws in the reference genome. We describe the concepts underlying MUM-based analysis, the software implementation, and its usage.

We show via simulation that the MUMdex aligner and alignment format are able to correctly detect and record genomic events. We characterize alignment performance and output file sizes for human whole genome data and compare with Bowtie 2 and the BAM format. Preliminary results show the practicality of the analysis approach by detecting de novo mutation candidates in human whole genome DNA sequence data from 510 families. We provide a population database of events from these families for use by others.

A Platform for Access and Analysis of Genetic Variants in Phenotype-Rich Family Collections

Y-h. Lee, B. Yamrom, S. Marks, I. Iossifov [in collaboration with M. Cokol, Axcella, Boston; A. Nenkova, University of Pennsylvania; L. Chorbadjiev, SeqPipe Ltd., Sofia, Bulgaria]

WES, a technique that enables the inexpensive identification of genetic variants in the gene-encoding regions of the genomes of thousands of people, is quickly transforming human genetics. Particularly successful are the numerous studies that used WES in large collections of families to study the genetic architectures of human disorders with strong detrimental effect in fecundity, including autism, intellectual disability, schizophrenia, epilepsy, and congenital heart disease. These studies identified large numbers of genetic variants segregating in the families or arising de novo in children, gathered detailed phenotypic measurements of the studied individuals, and used the complex data sets to develop models of genotype and phenotype relationships.

There is an enormous amount of work that needs to follow the early success in the genetics of such complex disorders to develop effective treatment and early diagnostic strategies. A variety of future research projects will study in detail the effects of hundreds of genetic variants and genes at molecular, cellular, and organismic levels. Such projects will greatly benefit from the accumulated family WES data sets, but their

large size and complex structure create a major obstacle for their efficient use. Here, we present the GPF (genotype and phenotype in families) system that manages such data sets and has an intuitive interface that makes it possible for the wider scientific community to benefit from the new collections.

RNA-Seq of SSC

We are finalizing our analysis of the whole-genome data from approximately 2400 of the SSC families. The major result of that effort is the estimate of the contribution of the de novo noncoding variants. Specifically, we observed a significantly increased rate of de novo intronic variants in affected children compared with their unaffected siblings when we restrict the rate observation to the autism genes previously implicated by WES. The increase in the rate is consistent with a contribution of ~5% of de novo intronic indels to the autism diagnosis in SSC (see the section De Novo Indels within Introns Contribute to ASD Incidence above). We do not observe a similar increase in the rate of de novo intronic substitutions, but it is expected that the size of the study is insufficient to detect that signal given the much higher rate of background noise for substitutions. Nevertheless, we also expect that de novo intronic substitutions have a contribution and guess that is likely of similar magnitude to the contribution of the de novo intronic indels. As others have reported an increased rate of de novo mutation in affected versus unaffected children within the control regions of the intergenic space, we expect that the contribution of noncoding de novo mutation is close to 15%, perhaps only slightly less than the contribution from de novo coding mutation.

Despite the large contribution of the noncoding variants, we have no good purely analytic method to distinguish the specific causal sequence variation from the many random ones. We proposed to address that through study of the RNA. We expect that for the majority of the causal de novo noncoding variants, the immediate effect would be on the expression of nearby genes, and such changes in expression can be detected through RNA-Seq by comparing the expression of the affected gene allele with that of the unaffected allele, a method called allele-specific expression (ASE). In the last year, we initiated pilot experiments to test the feasibility of this approach. This is a collaborative effort, including several groups at CSHL (Wigler and Levy labs), several groups at NYGC (New York Genome Center; Hemali Phatnani's and Tuuli Lappalainen's groups and Tom Maniatis), and Kristen Baldwin from the Scripps Research Institute.

We have access to Epstein–Barr virus (EBV) immortalized lymphoblastoids (LCLs) for all the individuals of the SSC. One of our pilots addressed the question of whether LCLs are a good source to study ASE. In collaboration with Kristen Baldwin, we transformed an LCL into an induced pluripotent stem cell (iPSC) further down neuronal lineages. We then generated and sequenced RNA libraries from the original LCLs, iPSCs, and derived neurons. The analysis of these data is ongoing, but we have already made a few useful observations. First, the transformation processed worked successfully. Second, the derived neurons express nearly 90% of the autism genes identified by exome sequencing, whereas the LCLs express ~70% of these genes. Third, when a gene is expressed in both cell types, the ASE is preserved.

In an additional pilot, we performed RNA-Seq from LCLs of six of the SSC families, two of which had identical affected twins and an unaffected sibling and four of which had one affected and one unaffected child. Among the goals of this pilot were tuning our bench protocols and analytical tools and estimating the noise sources (like *trans*-regulatory and epigenetic effects) that would decrease the power of detecting ASE. It appears from the pilot that such noise sources are not negligible, but are manageable.

PUBLICATIONS

Ye K, Iossifov I, Levy D, Yamrom B, Buja A, Krieger AM, Wigler M. 2017. Measuring shared variants in cohorts of discordant siblings with applications to autism. *Proc Natl Acad Sci* **114:** 7073–7076.

In Press

Buja A, Volfovsky N, Krieger AM, Lord C, Lash AE, Wigler M, Iossifov I. 2018. Damaging de novo mutations diminish motor skills in children on the autism spectrum. *Proc Natl Acad Sci* **115:** E1859–E1866.

SEQUENCE–FUNCTION RELATIONSHIPS AND OTHER QUANTITATIVE PROBLEMS IN MOLECULAR BIOLOGY

J.B. Kinney A. Ayaz W-C. Chen K. Coshic T. Forcier A. Tareen

During Dr. Kinney's graduate and postdoctoral training, he developed Sort-Seq, the first massively parallel reporter assay capable of being used in living cells. A single Sort-Seq experiment can measure the transcriptional activities of ~10^5 different bacterial promoters. A quantitative analysis of the resulting data can allow one to dissect the molecular mechanisms of promoter function (e.g., identify transcription factor–binding sites and RNA polymerase–binding sites), as well as quantitatively measure the in vivo interactions between such regulatory proteins.

Sort-Seq represents just one example of how ultra-high-throughput DNA sequencing can be used to discern quantitative models of complex biophysical/biochemical systems. The Kinney lab has since developed or helped to develop other massively parallel assays that can provide similar quantitative insight into other systems. The Kinney lab is also developing new mathematical and computational methods—in part for analyzing the data produced by these experimental techniques, but also for addressing more general and fundamental problems in statistical learning.

Measuring the Sequence-Affinity Landscape of Antibodies with Massively Parallel Titration Curves

Despite the central role that antibodies play in the adaptive immune system and biotechnology, surprisingly little is known about how antibody sequence quantitatively specifies antibody–antigen binding affinity. For instance, we cannot predict the number of different antibody sequences that can bind an antigen of interest with a given affinity, how this density of states varies with binding affinity, and so on.

In collaboration with Aleksandra Walczak and Thierry Mora (Ecole Normale Supérieure de Paris), the Kinney lab has developed a new experimental technique, called Tite-Seq, which is aimed at addressing such problems (Adams et al., *eLife 5:* e23156 [2016]). Tite-Seq combines yeast display and fluorescence-activated cell sorting (FACS) in a way that allows one to measure the affinity of thousands of variant antibody proteins in a single experiment (see Fig. 1). In contrast to existing deep mutational scanning assays, which measure binding at only a single ligand concentration, Tite-Seq provides full binding titration curves. Having these titration curves eliminates the confounding factors present in deep mutational scanning assays, such as the sequence dependence of antibody expression and stability. More recently, the Kinney lab has used these data to explore the quantitative properties of epistasis in antibody–antigen binding (Adams et al. 2017). Interestingly, they find that binding energy is, as a molecular phenotype, less epistatic than any other monotonic function of binding energy, such as affinity.

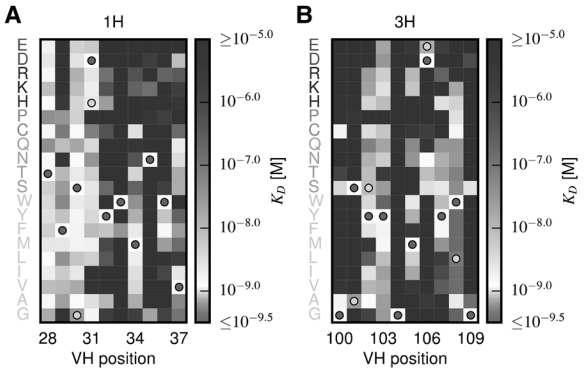


Figure 1. The sequence-affinity landscape of the (*A*) CDR1H and (*B*) CDR3H regions of the antifluorescein antibody studied in Adams et al. (*eLife 5:* e23156 [2016]) using Tite-Seq. The affinity resulting from each possible single-residue mutation within these two 10-residue regions is shown. Dark gray dots indicate the wild-type residue at each position. Tite-Seq is the first deep mutational scanning assay to provide absolute affinity measurements such as these.

Learning Quantitative Sequence–Function Relationships from Massively Parallel Assays

Prior work of Dr. Kinney's has shown that quantitative models of sequence–function relationships can be inferred from the data produced by Sort-Seq and other massively parallel assays. However, standard methods in statistical inference are not well suited for this modeling task. The reason is that standard inference methods rely on a quantity called "likelihood." To compute likelihood, one has to make strong assumptions about the quantitative form of experimental noise. The noise present in many massively parallel assays, however, is often difficult to precisely characterize.

Over the years, Dr. Kinney has published multiple papers showing how using a quantity called "mutual information" in place of likelihood can resolve this inference problem. Recently, Gurinder Atwal (CSHL) and Dr. Kinney reviewed the theoretical underpinnings of this problem (Atwal and Kinney, *J Stat Phys 162:* 1203 [2016]). They further introduced the mathematical concept of "dual modes" in the space of experimental noise models. Dual modes are closely related to the concept of diffeomorphic modes in parameter space, a concept that they introduced in a previous paper. To illustrate the difference between mutual-information-based and likelihood-based inferences, they also described the first analytically tractable model (the "Gaussian selection model") of inference from a massively parallel assay. In addition to their relevance to new biological experiments, these findings are applicable to a variety of other statistical problems.

Performing mutual-information-based analysis of massively parallel data remains a major practical challenge because of the lack of available software. To address this need, Bill Ireland and Dr. Kinney have created a software package, called MPAthic, which is described in our preprint (Ireland and Kinney, bioRxiv doi: 10.1101/054676 [2016]). MPAthic provides simple command line methods that can be used to fit a variety of quantitative models to a variety of massively parallel data sets using mutual information maximization, as well as other inference techniques.

Modeling Multiparticle Complexes in Stochastic Chemical Systems

Dr. Kinney's work on transcriptional regulation requires building quantitative biophysical models of molecular interactions. Currently, the only way to mathematically define models of such interactions is to explicitly list all the possible states of a system. This task, however, becomes exponentially more difficult as the number of molecular components increases. This exponential explosion in difficulty reflects a shortcoming of the mathematical tools that are currently used for studying such systems.

To address this problem, Muir Morrison (Caltech) and Dr. Kinney developed a mathematical formalism that represents stochastic chemical systems of multiparticle complexes in terms of component particles, interaction energies, and assembly rules. This mathematical formalism also has a diagrammatic representation that greatly aids in its use. In Morrison and Kinney (bioRxiv doi: 10.1101/045435 [2016]), they present this formalism and show how it can dramatically simplify the description of both equilibrium and nonequilibrium stochastic chemical systems. Their hope is that these methods will serve a purpose much like Feynman diagrams do in physics, bridging the gap between one's intuitive visual understanding of a system and rigorous quantitative models thereof.

PUBLICATIONS

Adams RM, Kinney JB, Walczak AM, Mora T. 2017. Physical epistatic landscape of antibody binding affinity. bioRxiv doi: 10.1101/232645.

Ipsaro JJ, Shen C, Eri A, Xu Y, Kinney JB, Joshua-Tor L, Vakoc CR, Shi J. 2017. Rapid generation of drug-resistance alleles at endogenous loci using CRISPR–Cas9 indel mutagenesis. *PLoS ONE* **12:** e0172177.

Senturk S, Shirole NH, Nowak DD, Corbo V, Pal D, Vaughan A, Tuveson DA, Trotman LC, Kepecs A, Kinney JB, Sordella R. 2017. Rapid and tunable method to temporally control gene editing based on conditional Cas9 stabilization. *Nat Commun* **8:** 14370.

In Press

Belliveau NM, Barnes SL, Ireland WT, Jones DL, Sweredoski MJ, Moradian A, Hess S, Kinney JB, Phillips R. 2018. A systematic approach for dissecting the molecular mechanisms of transcriptional regulation in bacteria. bioRxiv doi:10.1101/239335.

IN SILICO CANCER GENOMICS

A. Krasnitz P. Belleau A. Rudansky
N. Kostic J. Song

Research in our group is focused on in silico cancer genomics. In the last several years, there has been explosive growth in the volume as well as quality, variety, and detail of cancer-related genomic data available for analysis. This development came about through rapid accumulation of public data sets generated by individual laboratories worldwide and large-scale cancer genome projects such as The Cancer Genome Atlas (TCGA) and with the advent of new experimental methodologies, especially next-generation sequencing and single-cell genomics. We see our goal in channeling this flood of data into a number of clinically relevant applications. These include discovery of genomic markers for clinical outcome and molecular classification of cancer; elucidating the clonal structure of tumors and its relation to progression, invasion, metastasis, and response to treatment; and pinpointing and prioritizing targets for functional analysis. Our work is performed in close coordination with experimental studies performed by the Wigler, Tuveson, Spector, Stillman, and Fearon laboratories at CSHL.

Single-Cell Genomics as a Tool for Early Detection of Cancer

Metastasis is by far the most lethal aspect of cancer. Lifesaving and curative treatments of tumors are much more likely to succeed if performed before metastatic spread while the disease is still confined to the primary site. A major reduction of cancer mortality may therefore be achieved by early detection of tumors. To become part of clinical practice and make an impact on patient outcomes, an early detection method must be applicable across a wide range of tumor types and must be sensitive, cancer-specific, and affordable. All existing methods of early detection—whether based on blood-borne protein markers, such as prostate-specific antigen (PSA), or on imaging techniques, such as magnetic resonance imaging (MRI)—lack at least one of these attributes.

In response to this unmet need, we evaluated the potential for early detection of rare circulating tumor cells (CTCs) in the blood in asymptomatic patients, using sparse genomic profiling of single cells. A detection method we envision exploits a combination of two phenomena observed in neoplastic cells across multiple tumor types: clonal expansion and massive, genome-wide DNA copy number variation. We reasoned that CTCs representing a clonal population may be distinguished from other circulating cells by shared features of their DNA copy number profiles. To put this idea to a test, we used an in-house collection of single-cell genome sequences to simulate CTC-based detection of nearly 4000 tumors with published copy number profiles, spanning 11 major tumor types. We assumed that a minimum of 10 clonally related CTCs are present in a standard-sized blood draw from a patient. Using graph-theoretical methods of data analysis, we then showed that these cells are detectable with high success rate across multiple major tumor types, even if the CTCs are vastly outnumbered by nontumor cells in a blood-derived specimen.

Although the outcome of our computational study is encouraging, the success of the CTC-based approach to early detection of cancer ultimately depends on CTC concentration in blood at early stages of the disease. Because little reliable data exist for this crucial quantity, we recently initiated, in collaboration with clinical partners at Northwell Health, an assessment of CTC counts in the blood of patients newly diagnosed with ovarian cancer. For these patients, molecular signature of the tumor may be acquired from the tumor tissue specimen following surgery. We will then use a refined, error-free method of DNA amplification to accurately count cells bearing this signature in a blood specimen from the matching patient.

Diagnostic Molecular Pathology of Prostate Cancer

Recently, we used sparse single-nucleus sequencing and the computational pipeline (described in the following) to reconstruct the cell population structure

in samples from 11 sets of prostate biopsies, all with pathological grades, quantified as the Gleason score. The grades spanned a range of Gleason scores from benign to 9. Eight of the biopsies were performed diagnostically, with 10 to 15 biopsy cores per patient. Among these eight patients, five went on to undergo radical prostatectomy. Remarkably, for three of these five patients, pathological evaluation of the resected tissue differed from that of the diagnostic biopsy.

We set out to assess the potential of single-cell genomics in resolving such ambiguities. Approximately 30 single-cell genomes were acquired by sparse sequencing from each core and processed using our single-cell computational pipeline to yield integer genomic DNA copy number profiles for each cell. Genealogical trees formed by these genomes were derived for each patient, and clones were identified. There are several key observations from our analysis of tumor cell genealogies. The first is that clones are unlikely to form in nonmalignant prostates. In prostates in which the Gleason score indicated malignancy, cells in higher-grade regions display a well defined clonal identity and, within each clone, share a substantial number of genomic features. The second observation is that these clones mostly are formed by cells that share an anatomic location. At the same time, we do occasionally find evidence for clones spreading beyond their anatomic origins. Third, copy number profiles of cells in clonal populations grow progressively more complex with the Gleason scores, and the populations themselves grow more numerous in the prostatic tissue.

We introduced a number of simple clone-related genomic metrics to quantify these observations. For example, we defined clonal heterogeneity as the number of clones observed in the prostate. Importantly, we observed that such genomic measures are better correlated with the postsurgery Gleason score than conventional histopathological metrics at diagnosis, including the diagnostic Gleason score itself. These findings point to potential utility of single-cell genomic analysis to complement conventional pathology.

Computational Pipeline for Single-Cell Genomics

The study of genomic and transcriptional properties of individual cells is now the focus of multiple research laboratories around the world. In application to cancer, this line of research has revealed the genomic complexity of the disease and the presence of multiple genealogically related cell populations in a tumor. Detailed knowledge of the clonal structure of a cancer potentially is of high clinical value: Multiplicity of clones or lesions in most advanced clones is a possible measure of progression; spatial pattern of clone dispersal in a tumor may signal elevated propensity to invade; and lesions observed in individual clones but not in the bulk tissue may point to targets for therapy. DNA copy number profiling of cells from sparse sequencing is an accurate, economically feasible technological approach to the study of cancer subpopulation structure. Novel multiplex sequencing techniques, developed by the Wigler lab at CSHL, permit simultaneous sequencing of hundreds of single-cell DNA specimens and their subsequent copy number profiling at up to 50 kb resolution. Optimal use of this data form for robust reconstruction of cancer-cell phylogenies is a challenging computational problem calling for new and robust informatic and statistical tools.

We responded to this challenge by developing a computational pipeline for single-cell genomics. The pipeline comprises two major modules, one for deriving integer-valued copy number profiles of individual cells and one for establishing genealogical relations among the cells in a sample and identification of clones. The input into the pipeline consists of cell-specific sets of sequencing reads. These are first aligned with the genome. The sequencing read density as a function of genomic position is then used to derive integer-valued DNA copy number profiles for each cell. Before any further processing these profiles are examined for evidence of extensive DNA damage or degradation, and profiles derived from damaged DNA are filtered out. Next, each of the remaining profiles is reduced to a set of copy number change points in which, for each change point, the genomic interval of its likely location and the sign of the change are specified.

From this point on we combine change-point reduced copy number profiles of the cells constituting the sample with the ultimate goal of reconstructing their phylogeny. To this end, we first derive a minimal set of features to account for all the change points observed in the sample and construct a table to indicate, for each cell genome, the presence or absence of each feature. Pairwise similarity of cell genomes is quantified based on the number of features shared

by the pair. We then examine the resultant similarity matrix for the presence of clones—that is, groups of cells whose unusually high number of shared features can only be explained by their descent from a recent common ancestor. Such clones may in turn contain subclones of even more closely related cells.

Single-cell data as interpreted by the pipeline must be presented to cancer biologists and clinicians in a comprehensible way to have impact on the clinical outcome. In practice, this means that the data should be viewable in an organized way with the viewer retaining the ability to change the organization. The end user must be able to see and navigate the phylogenetic organization. Moreover, the data and their interpretation must be viewed in the context of other parameters such as anatomical sites along with their pathological assessment. Such examination of the data in aggregate offers the best chance to reveal the critical properties of the sample relevant to clinical assessment. An integrated user interface capable of handling specialized forms of data that arise in single-cell research is therefore essential. With this necessity in mind, we have created and made public the Single-Cell Genome Viewer (SCGV). The viewer displays multiple single-cell genomic profiles in the chromosomal order with colors encoding the sign and magnitude of copy number aberrations. The cells on display are ordered as leaves on a tree, reflecting the structure of the population being sampled. In particular, clonal and subclonal identities of the cells are indicated in the appropriate tracks. Other tracks provide information on the cell ploidy, the quality of the cell genomic data, and, importantly, the anatomic origin of the cell. Elements of the data can be examined in greater detail by a combination of selection and zoom-in. For example, any number of cells can be selected and the corresponding copy number profiles plotted in a separate interface. A link to the University of California, Santa Cruz (UCSC) genome browser is available for any genomic region displayed by the viewer to help put the findings in a broader genomic context.

PanCanAtlas: Stemness and Immune Response across Multiple Tumor Types

TCGA is a massive, comprehensive data repository for cancer research. More than 30 cancer types are currently represented in the atlas with hundreds of patient cases per type. For each tumor somatic mutation, DNA copy number variation, DNA methylation, mRNA and micro-RNA profiles are available along with clinical annotation for the patient. By combining multiple data sets, TCGA consortium generated integrated molecular portraits of a number of cancer types. PanCanAtlas is a successor project to TCGA, seeking to take data integration one step further and across multiple cancer types. Our group has been part of this effort since June 2015, with two focal areas of interest. The first of these is concerned with quantifying the presence of stem-cell-like cell populations in tumors. In particular, loss of differentiation has long been recognized as a key feature of cancer pathology. We hypothesize that, on a cellular level, this loss of differentiation is accompanied by acquisition of stem-cell-like properties. In an exploratory study, we find that genes associated with loss of differentiation in multiple cancers frequently occur in published transcriptional signatures of embryonic stem cells.

In addition, we are interested in pan-cancer characterization of immune response to tumors with a particular focus on the role of cancer testis antigens (CTAs)—that is, products of genes that are exclusively expressed in testes of healthy male individuals where they are protected by immune privilege from exposure to the immune system. Expression of these genes in tumors should therefore elicit an immune response, in particular, by cytotoxic T cells. Such a response is an end result of a multistep process involving mRNA expression and translation into proteins, followed by proteasome cleavage, antigen presentation, and recognition by the adaptive immune system. Our ongoing study aims at taking these multiple steps into account to quantify potential immune response to CTAs in tumors documented by TCGA. With this ultimate goal in mind, we have collaborated with Seven Bridges Genomics to determine HLA class I A, B, and C alleles for nearly 9000 TCGA patients with available RNA-sequencing data. These results were used as input for multiple additional projects—most prominently, identification of neoantigens for all patients in this massive cohort.

PUBLICATIONS

Krasnitz A, Kendall J, Alexander J, Levy D, Wigler M. 2017. Early detection of cancer in blood using single-cell analysis: A proposal. *Trends Mol Med* **23:** 594–603.

Alexander J, Kendall J, McIndoo J, Rodgers L, Aboukhalil R, Levy D, Stepansky A, Sun G, Chorbadjiev L, Riggs M, et al. 2017. Utility of single cell genomics in diagnostic evaluation of prostate cancer. *Cancer Res* **78:** 348–358.

Chorbadjiev L, Kendall J, Zhygulin V, Alexander J, Wigler M, Krasnitz A. 2017. KrasnitzLab/SCGV: SCGV is an interactive graphical tool for single-cell genomics data, with emphasis on single-cell genomics of cancer. Computer software, version 1.0.1. https://zenodo.org/record/1095780#.WimrDd9fifY

Kato M, Vasco DA, Sugino R, Narushima D, Krasnitz A. 2017. Sweepstake evolution of cancer genomes revealed by population-genetic analysis of copy-number alterations in single-cell genomes of breast cancer. *R Soc Open Sci* **4:** 171060.

COMPUTATIONAL GENETICS

D. Levy A. Moffitt M. Wroten

The data-rich environment at CSHL generates a wealth of opportunities for the application of mathematics and computation to further our understanding of biology and genetics. The primary activities of our laboratory are algorithm and protocol development, data analysis, and genetic theory, with a focus on sporadic human diseases such as autism, congenital heart disease, and cancer.

Quantitative-Sensitive Detection

Present-day Illumina sequencing technology generates hundreds of millions of high-quality sequence reads with error rates in the range of 1 per 100 bases. For most purposes, these error rates are adequate. However, there are a class of applications that require the measurement of rates lower than that. These include important applications such as measuring residual disease in cancer, measuring subpopulations within a tumor, and detecting somatic and germline variation.

Together with Michael Wigler and Zihua Wang, Andrea Moffitt and I developed a protocol that adds a unique sequence identifier (or "varietal tag") to the initial template molecule. Because sequence error is sparse and independent of the template molecule, reads with the same varietal tag are unlikely to have the same sequencer errors. Taking a sequence consensus from all reads with the same tag corrects for sparse error. We found that systematic errors have a sequence-context specific signature and, by analyzing multiple high-depth tagged experiments, we are able to model the error of consensus reads. Some sequence contexts (such as CpG) have a high background rate that has nothing to do with machine error. However, other contexts are very stable, with background error rates for consensus sequences lower than one part in a million. The MASQ (multiplex accurate sensitive quantitation) protocol enables the simultaneous measurement of variants up to 50 loci and millions of templates per locus. We developed an informatics pipeline that selects variants with low background error rates, determined the optimal reagents for the protocol, and designed the necessary sequencing primers.

With Mona Spector, Alex Krasnitz, and Joan Alexander, we have applied this method to measure residual disease in patients treated for AML. Using MASQ, we measured tumor-specific variants in blood samples taken after treatment. In some patients, the residual load was detectable at levels observable by traditional sequencing. In other patients, we were able to measure rates as low as one part in 100,000. We are presently testing the utility of MASQ in the context of solid tumors.

Genomic Rearrangements

Together with Peter Andrews, Michael Wigler, and Ivan Iossifov, we developed an algorithm called MUMdex for identifying large genomic rearrangements—deletions, insertions, translocations, and inversions—by cataloging discontinuities in the mapping of reads against the reference genome. Taking advantage of memory-intensive computational methods, we are able to rapidly identify all maximal unique matches (MUMs) between a read and a reference genome. This provides a scaffold for cataloging and indexing all of the sequence data such that we can quickly identify discontinuities—places where placing the reads on the scaffold bends or breaks the reference genome.

We developed a calculus for recording and storing these discontinuities that satisfies algebraic properties that reveal aspects of the genomic event. By looking across all samples, we can identify common events, spurious rearrangements, and global discrepancies within the reference. With all reads indexed and anchored to the reference genome, we reassemble rare and de novo discontinuities within an individual or family. Analyzing large data sets requires an efficient and robust data format and processing algorithms, which we are currently deploying on thousands of whole-genome sequences.

Mutational Sequencing

Despite vast improvements in DNA sequencing, many problems of interpretation arise when trying to count or assemble molecules (templates) that are largely identical. A few years ago, we showed that, in theory, randomly mutating DNA templates before amplification resolves many of these problems. We can accurately count the number of templates by counting the number of unique patterns. By introducing distinctive patterns onto otherwise identical spans, we enhance our ability to correctly assemble sequences. This idea has applications in RNA expression analysis, haplotype phasing, copy number determination, and genome assembly. Mutational sequencing (muSeq) solves counting problems and effectively generates long reads from short-read sequence data.

This year, we published the first working protocol for muSeq using incomplete sodium bisulfite conversion as the mechanism for introducing mutations. Loosely speaking, sodium bisulfite converts a C to T in a DNA template and, by tuning to a conversion rate of 50%, we can label each template molecule with a unique mutational signature of C-to-T conversions. Clustering reads with the same conversion pattern enables accurate count and long-range assembly of initial template molecules from short-read sequence data. Using a PstI representation, we demonstrated that muSeq improves copy number measurement and significantly reduces sporadic sequencing error. Using a cDNA library, we demonstrated long-range assembly of template molecules up to 4 kilobases (kB) in length.

Presently, together with Siran Li, we are developing "targeted muSeq": a protocol for mutational sequencing applicable to any genomic region of <20 kB in length. The protocol is easy, inexpensive, and reproducible. Crucially, the informatics do not assume a prior genome performing a de novo assembly of each haplotype. We are testing and validating our protocol and algorithms on well characterized genomic regions with the aim of resolving haplotype-specific assembly for difficult-to-sequence genomic regions such as the HLA locus and the protocadherin genes. With minor modifications, we will extend our de novo assembly approach to cDNA, enabling isoform resolution transcriptome data in the absence of a reference genome.

PUBLICATIONS

Krasnitz A, Kendall J, Alexander J, Levy D, Wigler M. 2017. Early detection of cancer in blood using single-cell analysis: a proposal. *Trends Mol Med* **23:** 694–703.

Kumar V, Rosenbaum R, Wang Z, Forcier T, Ronemus M, Wigler M, Levy D. 2017. Partial bisulfite conversion for unique template sequencing. *Nucleic Acids Res* **46:** e10. doi:10.1093/nar/gkx1054

Ye K, Iossifov I, Levy D, Yamrom B, Buja A, Krieger AM, Wigler M. 2017. Measuring shared variants in cohorts of discordant siblings with applications to autism. *Proc Natl Acad Sci* **114:** 7073–7076.

PREDICTING EFFECTS OF MUTATIONS FROM HIGH-THROUGHPUT DATA

D. McCandlish J. Zhou

Recent developments in high-throughput mutagenesis and phenotyping have made it increasingly routine to simultaneously measure, in a single protein, the effects of tens of thousands of single and multiple mutations. In the McCandlish lab, we develop new computational and statistical tools to analyze and understand the results of these large experiments. Such experiments have revealed complex patterns of genetic interaction in which the same mutation can have very different effects depending on the rest of the genetic sequence in which it occurs. A key challenge in the group is to produce methods that can make sense of these intrinsically high-dimensional interactions and root them in basic biophysical principles. We are also interested in understanding the forces that cause protein sequences to change over evolutionary time, and we work to construct and analyze mathematical models that can account for observed patterns of variation in molecular sequences.

Probabilistic Models of Mutational Effects

Predicting the effects of amino acid replacements is important for both assessing the pathogenicity of missense variants in human genome sequences and understanding the dynamics of protein sequence change over evolutionary time. Comprehensive, high-throughput mutagenesis experiments measuring the effect of every possible single amino acid replacement have now been conducted on many model proteins. These experiments have the potential to transform our ability to predict the effects of mutations by providing a quantitative and unbiased view on the functional impact of amino acid replacements. However, the proteins that have been assayed using these techniques have very different functions and indeed are drawn from many different organisms, so that it is unclear how to assemble these results into a single synthetic model. Together with Arlin Stoltzfus at the National Institute of Standards and Technology, we have developed a probabilistic modeling framework based on nonparametric statistics to address this problem. For any given class of mutation, our model produces a probability distribution describing how deleterious this mutation is likely to be compared with other mutations in the same protein. Currently, we have sufficient data to infer these distributions for every possible pair of wild-type and mutant amino acids. Although many of these amino acid exchanges show highly concentrated distributions, indicating high confidence in our prediction, many others are nearly uniform, indicating that whether the mutation is likely to be deleterious or benign depends primarily on the sequence context and not merely on the similarities between the wild-type and mutant amino acid. Surprisingly, we find that, essentially, all of the inferred distributions are well approximated by a simple one-parameter family of probability distributions, which will allow these distributions to be efficiently incorporated into a variety of downstream bioinformatics applications.

Modeling Genetic Interactions

The effects of a mutation may depend on which other mutations are present in the genetic sequence, a phenomenon known as epistasis. Although interactions between a single pair of mutations are relatively easy to understand, it has become increasingly clear that functional interactions in proteins may involve many different amino acid sites. Currently available methods for modeling the relationship between protein sequence and function are capable of incorporating pairwise interactions, but there is a need for new methods that can accommodate "higher-order" genetic interactions between more than two sites.

Inspired by models from protein biophysics that consider the thermodynamic effects of mutations on the fraction of time a protein is folded, we are collaborating with Jakub Otwinowski and Joshua Plotkin at the University of Pennsylvania to develop models that incorporate global interactions between all sites in a protein. Similar to how the fraction of folded protein can be modeled as a nonlinear function of a free energy of folding that is approximately additive

between sites, our model assumes that the phenotype observed in an experiment is determined by a nonlinear mapping from an underlying additive trait, and we simultaneously infer the form of the mapping and the coefficients of the underlying additive model. Using publically accessible data from high-throughput mutagenesis experiments, we have shown that our model can capture essentially all of the genetic interactions observed above the experimental noise floor for two model proteins, GFP and protein G.

Although the above model successfully incorporates some forms of higher-order genetic interaction, some of the most interesting forms of genetic interaction, such as mutations that are advantageous on one background but disadvantageous on another, cannot be accommodated in this framework. Synthesizing ideas from the contemporary engineering literature on signal processing with classical analyses of genetic interactions from the evolutionary biology literature, we have developed a set of statistical techniques that can fit arbitrarily complex forms of genetic interaction while nonetheless behaving in a simple and easily comprehensible manner in which data are sparse or absent. We have been using this method to impute missing data from combinatorial mutagenesis experiments that attempt to produce all possible combinations of amino acids at up to four amino acid sites, and we are currently working to develop efficient algorithms to extend this method to whole-domain and whole-protein mutagenesis.

Molecular Evolution

Besides our core interest in developing methods to predict the effects of mutations using data from high-throughput assays, we have also worked to address fundamental issues in molecular evolution. One important issue concerns the role of mutational biases in adaptive evolution. Classical arguments suggest that mutational biases should have only a minor effect on adaptive evolution because mutation rates are very small compared with the differences in reproductive rates between genotypic variants. However, more recent arguments have suggested that under some circumstances mutation rates can have a major impact on the direction of adaptation. In joint work with Arlin Stoltzfus from the National Institute of Standards and Technology, we attempted to adjudicate this controversy by considering documented cases in which the same adaptive, function-altering amino acid substitution has gone to fixation in multiple independent populations. When we looked at the nucleotide changes responsible for these amino acid substitutions, we found that these mutations were highly enriched for transitions over transversions, consistent with an effect of mutational bias on adaptation.

Another important issue concerns evolutionary dynamics within and between populations. In collaboration with Francois Bienvenu and Stephane Legendre at the École Normale Supérieure in Paris, as well as Erol Ackay at the University of Pennsylvania, we conducted a mathematical investigation of a class of deterministic models that are frequently used in ecology and evolution to model demography within populations, as well as evolution in populations with high mutation rates as occur in viruses and tumors. We developed a formal decomposition of these models and used this decomposition to show that many statistics commonly used to understand these models depend only on certain very specific characteristics of the genealogies of individuals within the population.

Finally, it has long been observed that mutations do not accumulate within populations evenly over time, but instead tend to occur in clusters, a phenomenon known as overdispersion of the molecular clock. We have developed a mathematical theory for the clustering of events in molecular evolution, and have shown that it can be used to approximate waiting times for certain events to occur such as the waiting time for a population to first produce a particular phenotype by mutation.

PUBLICATIONS

Bienvenu F, Akcay E, Legendre S, McCandlish DM. 2017. The genealogical decomposition of a matrix population model with applications to the aggregation of stages. *Theor Popul Biol* **155:** 69–80.

Stoltzfus A, McCandlish DM. 2017. Mutational biases influence parallel adaptation. *Mol Biol Evol* **34:** 2163–2172.

COMPUTATIONAL SEQUENCE ANALYSIS

M. Schatz H. Fang S. Ramakrishnan
M. Nattestad F. Sedlazeck

Our lab develops novel computational analysis methods to study the structure and function of genomes, especially genomes with medical or agricultural significance. These include methods for assembling sequence data into complete genomes, aligning sequences to discover variations or measure transcription levels, and methods for mining these data to discover relationships between an organism's genome and its traits. We then apply these methods to study several biological systems, such as to discover mutations associated with human diseases such as cancer and autism or to reconstruct the genomes of important agricultural crops and biofuels.

In the past year, we published several significant methods for analyzing biological sequencing data. In collaboration with the Tuveson laboratory, we developed GECCO, the Genomic Enrichment Computational Clustering Operation, which we used to study noncoding regulatory mutations in pancreatic cancer. With it, we identified many previously unknown mutations, which we could associate with substantial changes in gene expression in the cancer—including some that we could associate with reduced survival outcomes for the patients. In related work, we collaborated with the Sordella laboratory to study tumor heterogeneity and found that $CD44^{+}/CD24^{-}$ cancer cells are genetically highly unstable because of intrinsic defects in their DNA-repair capabilities. Another major advance was the publication of GenomeScope, a new statistical approach to infer the overall properties of genomes, such as their overall size, heterozygosity, or repeat content from unassembled short-read sequencing. This approach is especially important for studying nonmodel genomes in which these basic genome properties may be unknown. In collaboration with the McCombie laboratory, we also published the computational tools Alpaca and ECTools for the hybrid error correction and de novo assembly of long-read sequencing data. These tools can be used when a limited amount of long-read sequencing data is available from the sample, and we apply it to study the genomes of rice and the model legume *Medicago truncatula*. A final major new method was the publication of Scikit-ribo in collaboration with the Lyon and Siepel laboratories. Scikit-ribo is a new quantitative approach for analyzing Ribo-seq data, which can be used for studying how mRNA is translated into proteins using high-throughput sequencing. With it, we improve on previous analyses of these data to more accurately infer the true abundances of different proteins in yeast and *Escherichia coli*, thus laying the foundation for future studies in many other species. Other major works include coediting a special issue of the *Proceedings of the IEEE* on "The Bioinformatics of DNA" and a research highlight on methylation analysis using Oxford Nanopore sequencing published in *Nature Methods*.

Lab member Han Fang defended his thesis entitled "Graphical and machine learning algorithms for large-scale genomics data." For this work he received the prestigious President's Award to Distinguished Doctoral Students at Stony Brook University. After graduation, Dr. Fang joined the company Facebook as a data scientist so that he can apply his quantitative skills to even larger data sets. Lab member Maria Nattestad also defended her thesis "Computational methods for analysis and visualization of long-read sequencing data in cancer genomics" in the Watson School of Biological Sciences at CSHL. She accepted a position as Scientific Visualization Lead at the genomics company DNAnexus, where she is now leading a team of scientists and engineers to develop new ways to visualize and understand large-scale sequencing data. Postdoctoral fellow Fritz Sedlazeck also transitioned to the Baylor College of Medicine, where he will soon start his own computational research laboratory.

Recurrent Noncoding Regulatory Mutations in Pancreatic Ductal Adenocarcinoma

The contributions of coding mutations to tumorigenesis are relatively well known; however, little is known about somatic alterations in noncoding DNA. Here,

we describe GECCO to analyze somatic noncoding alterations in 308 pancreatic ductal adenocarcinomas (PDAs) and identify commonly mutated regulatory regions. We find recurrent noncoding mutations to be enriched in PDA pathways, including axon guidance and cell adhesion, and newly identified processes, including transcription and homeobox genes. We identified mutations in protein binding sites correlating with differential expression of proximal genes and experimentally validated effects of mutations on expression. We developed an expression modulation score that quantifies the strength of gene regulation imposed by each class of regulatory elements and found the strongest elements were most frequently mutated, suggesting a selective advantage. Our detailed single-cancer analysis of noncoding alterations identifies regulatory mutations as candidates for diagnostic and prognostic markers, and suggests new mechanisms for tumor evolution.

TGF-β Reduces DNA Double-Strand Break Repair Mechanisms to Heighten Genetic Diversity and Adaptability of $CD44^+/CD24^-$ Cancer Cells

Many lines of evidence have indicated that both genetic and nongenetic determinants can contribute to intratumor heterogeneity and influence cancer outcomes. Among the best described subpopulations of cancer cells generated by nongenetic mechanisms are cells characterized by a $CD44^+/CD24^-$ cell surface marker profile. Here, we report that human $CD44^+/CD24^-$ cancer cells are genetically highly unstable because of intrinsic defects in their DNA-repair capabilities. In fact, in $CD44^+/CD24^-$ cells, constitutive activation of the TGF-β axis was both necessary and sufficient to reduce the expression of genes that are crucial in coordinating DNA damage repair mechanisms. Consequently, we observed that cancer cells that reside in a $CD44^+/CD24^-$ state are characterized by increased accumulation of DNA copy number alterations, greater genetic diversity, and improved adaptability to drug treatment. Together, these data suggest that the transition into a $CD44^+/CD24^-$ cell state can promote intratumor genetic heterogeneity, spur tumor evolution, and increase tumor fitness.

16GT: A Fast and Sensitive Variant Caller Using a 16-Genotype Probabilistic Model

16GT is a variant caller for Illumina whole-genome and whole-exome sequencing data. It uses a new 16-genotype probabilistic model to unify single-nucleotide polymorphism and insertion and deletion calling in a single-variant calling algorithm. In benchmark comparisons with five other widely used variant callers on a modern 36-core server, 16GT showed improved sensitivity in calling single-nucleotide polymorphisms, and it provided comparable sensitivity and accuracy for calling insertions and deletions as compared with the GATK HaplotypeCaller. 16GT is available at https://github.com/aquaskyline/16GT.

GenomeScope: Fast Reference-Free Genome Profiling from Short Reads

GenomeScope is an open-source web tool to rapidly estimate the overall characteristics of a genome, including genome size, heterozygosity rate, and repeat content from unprocessed short reads. These features are essential for studying genome evolution and help to choose parameters for downstream analysis. We show its accuracy on 324 simulated and 16 real data sets with a wide range in genome sizes, heterozygosity levels, and error rates. GenomeScope is available at http://genomescope.org.

Reference Quality Assembly of the 3.5-Gb Genome of *Capsicum annuum* from a Single Linked-Read Library

Linked-read sequencing technology has recently been used successfully for de novo assembly of human genomes; however, the utility of this technology for complex plant genomes is unproven. We evaluated the technology for this purpose by sequencing the 3.5-gigabase (Gb) diploid pepper (*Capsicum annuum*) genome with a single linked-read library. Plant genomes, including pepper, are characterized by long, highly similar repetitive sequences. Accordingly, significant effort is used to ensure that the sequenced plant is highly homozygous

and the resulting assembly is a haploid consensus. With a phased assembly approach, we targeted a heterozygous F_1 derived from a wide cross to assess the ability to derive both haplotypes and characterize a pungency gene with a large insertion/deletion. The Supernova software generated a highly ordered, more contiguous sequence assembly than all currently available *C. annuum* reference genomes. More than 83% of the final assembly was anchored and oriented using four publicly available de novo linkage maps. A comparison of the annotation of conserved eukaryotic genes indicated the completeness of assembly. The validity of the phased assembly is further shown with the complete recovery of both 2.5-kb insertion/deletion haplotypes of the PUN1 locus in the F_1 sample, which represents pungent and nonpungent peppers, as well as nearly full recovery of the BUSCO2 gene set within each of the two haplotypes. The most contiguous pepper genome assembly to date has been generated and shows that linked-read library technology provides a tool to de novo assemble complex highly repetitive heterozygous plant genomes. This technology can provide an opportunity to cost-effectively develop high-quality genome assemblies for other complex plants and compare structural and gene differences through accurate haplotype reconstruction.

LRSim: A Linked-Read Simulator Generating Insights for Better Genome Partitioning

Linked-read sequencing, using highly multiplexed genome partitioning and barcoding, can span hundreds of kilobases to improve de novo assembly, haplotype phasing, and other applications. Based on our analysis of 14 data sets, we introduce LRSim, which simulates linked reads by emulating the library preparation and sequencing process with fine control over variants, linked-read characteristics, and the short-read profile. We conclude from the phasing and assembly of multiple data sets, recommendations on coverage, fragment length, and partitioning when sequencing genomes of different sizes and complexities. These optimizations improve results by orders of magnitude and enable the development of novel methods. LRSim is available at https://github.com/aquaskyline/LRSIM.

Hybrid Assembly with Long and Short Reads Improves Discovery of Gene Family Expansions

Long-read and short-read sequencing technologies offer competing advantages for eukaryotic genome-sequencing projects. Combinations of both may be appropriate for surveys of within-species genomic variation. We developed a hybrid assembly pipeline called "Alpaca," which can operate on 20× long-read coverage plus about 50× short-insert and 50× long-insert short-read coverage. To preclude collapse of tandem repeats, Alpaca relies on base-call-corrected long reads for contig formation.

Compared with two other assembly protocols, Alpaca showed the most reference agreement and repeat capture on the rice genome. On three accessions of the model legume *M. truncatula*, Alpaca generated the most agreement to a conspecific reference and predicted tandemly repeated genes absent from the other assemblies. Our results suggest Alpaca is a useful tool for investigating structural and copy number variation within de novo assemblies of sampled populations.

Scikit-ribo Enables Accurate Estimation and Robust Modeling of Translation Dynamics at Codon Resolution

Ribosome profiling (Ribo-seq) is a powerful technique for measuring protein translation; however, sampling errors and biological biases are prevalent and poorly understood. Addressing these issues, we present Scikit-ribo (https://github.com/schatzlab/scikit-ribo), an open-source analysis package for accurate genome-wide A-site prediction and translation efficiency (TE) estimation from Ribo-seq and RNA-sequencing data. Scikit-ribo accurately identifies A-site locations and reproduces codon elongation rates using several digestion protocols ($r = 0.99$). Next, we show that the commonly used reads per kilobase of transcript per million mapped reads–derived TE estimation is prone to biases, especially for low-abundance genes. Scikit-ribo introduces a codon-level generalized linear model with ridge penalty that correctly estimates TE while accommodating variable codon elongation rates and mRNA secondary structure. This corrects the TE errors for more than 2000 genes in *Saccharomyces*

cerevisiae, which we validate using mass spectrometry of protein abundances ($r = 0.81$), and allows us to determine the Kozak-like sequence directly from Ribo-seq. We conclude with an analysis of coverage requirements needed for robust codon-level analysis and quantify the artifacts that can arise from cycloheximide treatment.

PUBLICATIONS

Fang H, Huang Y, Radhakrishnan A, Siepel A, Lyon GJ, Schatz MC. 2017. Scikit-ribo enables accurate estimation and robust modeling of translation dynamics at codon resolution. *Cell Syst* **6:** 180–191.

Feigin ME, Garvin T, Bailey P, Waddell N, Chang DK, Kelley DR, Shuai S, Gallinger S, McPherson JD, Grimmond SM, et al. 2017. Recurrent noncoding regulatory mutations in pancreatic ductal adenocarcinoma. *Nat Genet* **49:** 825–833.

Heath LS, Bravo HC, Caccamo M, Schatz MC. 2017. Bioinformatics of DNA. *Proc IEEE* **105:** 419–421.

Luo R, Schatz MC, Salzberg SL. 2017. 16GT: A fast and sensitive variant caller using a 16-genotype probabilistic model. *Gigascience* **6:** 1–4.

Luo R, Sedlazeck FJ, Darby CA, Kelly SM, Schatz MC. 2017. LRSim: A linked reads simulator generating insights for better genome partitioning. *Comput Struct Biotechnol J* **15:** 478–484.

Miller JR, Zhou P, Mudge J, Gurtowski J, Lee H, Ramaraj T, Walenz BP, Liu J, Stupar RM, Denny R, et al. 2017. Hybrid assembly with long and short reads improves discovery of gene family expansions. *BMC Genomics* **18:** 541.

Pal, D Pertot A, Shirole NH, Yao Z, Anaparthy N, Garvin T, Cox H, Chang K, Rollins F, Kendall J, et al. 2017. TGF-β reduces DNA ds-break repair mechanisms to heighten genetic diversity and adaptability of $CD44^+/CD24^-$ cancer cells. *eLife* doi: http://dx.doi.org/10.7554/eLife.21615.

Schatz MC. 2017. Nanopore sequencing meets epigenetics. *Nat Methods* **14:** 347–348.

Vurture GW, Sedlazeck FJ, Nattestad M, Underwood CJ, Fang H, Gurtowski J, Schatz MC. 2017. GenomeScope: Fast reference-free genome profiling from short reads. *Bioinformatics* **33:** 2202–2204.

In Press

Hulse-Kemp AM, Maheshwari S, Stoffel K, Hill TA, Jaffe D, Williams S, Weisenfeld N, Ramakrishnan S, Kumar V, Shah P, et al. 2018. Reference quality assembly of the 3.5-Gb genome of *Capsicum annuum* from a single linked-read library. *Hortic Res* doi: 10.1038/s41438-017-0011-0.

POPULATION GENETICS AND TRANSCRIPTIONAL REGULATION

A. Siepel A. Blumberg, N. Dukler, B. Gulko, H. Hijazi, Y. Huang, M. Hubisz, E. Hutton, L. Pipes, R. Ramani, Y. Zhao

For the past several years, our research has focused on two major areas: human population genetics and transcriptional regulation in humans and *Drosophila*. The research in population genetics is performed either with publicly available genomic sequence data or with a variety of collaborators, whereas most of the work on transcriptional regulation is performed with our collaborators Charles Danko and John Lis at Cornell. We also have smaller collaborative projects on topics ranging from comparative transcriptomics of primates (in collaboration with Chris Mason, Weill Cornell Medical College), to molecular evolution of micro-RNAs in *Drosophila* (in collaboration with Eric Lai, Memorial Sloan Kettering Cancer Center), to prediction of the fitness consequences of mutations in rice, maize, and other crops (in collaboration with Michael Purugganan, NYU and Ed Buckler, Cornell), to studying the speciation process of recently diverged *Sporophila* songbirds (in collaboration with John "Irby" Lovette, Cornell, and Ilan Gronau, Herzliya Interdisciplinary Center, Israel), to combined experimental and computational characterization of *cis*-regulatory sequences in the human genome (in collaboration with Barak Cohen, Washington University, St. Louis). We focus on theoretical and computational research and do not generate our own data, but we often work closely with experimental collaborators on projects that have substantial experimental as well as computational components. We are broadly interested in molecular evolution, population genetics, and gene regulation, as well as machine learning, probabilistic modeling, and Bayesian statistics, and our research projects cut a broad swath across these diverse areas. Our research group is highly interdisciplinary, with members trained in computer science, mathematics, physics, genetics, and biochemistry, among other areas. The size of the group is stable at present, but there has been some recent turnover, with two Ph.D. students graduating in mid-2017 (Lenore Pipes and Brad Gulko) and two new postdoctoral associates starting in early 2018 (Hussein Hijazi and Yixin Zhao).

Below we describe recent progress in three main research areas.

Reconstruction of Demographic History from Complete Genome Sequences

Several years ago, we developed a statistical method based on the theoretical framework of the coalescent for reconstructing the demographic history of complex, structured populations from DNA sequence data. Our method, called *G-PhoCS* (Generalized Phylogenetic Coalescent Sampler), uses Markov chain Monte Carlo techniques to explore coalescent genealogies consistent with a particular population phylogeny, allowing for gene flow between designated populations. *G-PhoCS* produces Bayesian estimates of the key parameters that define these population phylogenies, such as the divergence times between populations, effective sizes of ancestral populations, and rates of postdivergence gene flow. We originally used the method to estimate the date of origin of one of the earliest branching extant human populations, the Khoisan hunter–gatherers of southern Africa. More recently, we have used it in collaboration with other research groups to shed light on the demographic histories of dogs and wild canids, and birds from the genus *Sporophila* (e.g., Campagna et al., *Mol Ecol 24:* 4238 [2015]).

G-PhoCS and methods like it "cheat" by considering only short, widely spaced genomic sequences and ignoring the difficult problem of modeling recombination. However, it would be preferable to consider not only the process of coalescence (finding common ancestry) at each locus in a genome, but also the manner in which historical recombination events alter these genealogies along the genome sequence. This combined history of coalescence and recombination can be explicitly represented by a generalized representation known as an "ancestral recombination graph,"

or ARG. The problem of reconstructing an ARG from sequence data, however, is notoriously difficult, and ARG inference has not been widely used in applied population genomics. Recently, we developed an algorithm for sampling ARGs within an approximate framework known as the sequentially Markov coalescent (SMC). Our method, called *ARGweaver*, uses techniques from hidden Markov models to repeatedly "thread" individual sequences through an ARG, leading to a Gibbs sampler over the space of ARGs. *ARGweaver* is the first ARG inference method that is efficient enough to apply to complete mammalian genomes. We have shown that it works remarkably well on simulated data and reveals clear signatures of natural selection in real human genome sequences.

In collaboration with S. Castellano, S. Pääbo, and colleagues (Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany), we recently applied *G-PhoCS* and *ARGweaver* in a joint analysis of the genome sequences of three Neanderthals, a Denisovan, and representatives of modern African, European, and Asian populations (Kuhlwilm et al., *Nature 530:* 429 [2016]). Using *G-PhoCS*, we were able to validate several previous findings of gene flow between archaic and modern humans. To our surprise, our joint modeling framework also allowed us to detect significant evidence of gene flow from modern humans into the Altai Neanderthal genome sequence in the opposite direction and much earlier than the previously reported human/Neanderthal introgression event. An analysis with *ARGweaver* supported this finding and allowed us to date the human-to-Neanderthal introgression event at ~100,000 years ago. Interestingly, this event substantially predates current estimates of the main out-of-Africa migration, and therefore suggests an earlier migration of modern humans out of Africa.

Our work on Neanderthal–human interbreeding suggests that a major advantage of *ARGweaver* is that it is especially powerful for identifying very early introgression events. However, a limitation of the method, at present, is that it naïvely assumes a prior distribution based on a single randomly mating population of constant size. Melissa Hubisz in the group is therefore working to extend *ARGweaver* to consider a full demographic model consisting of a tree of related populations of different sizes with gene flow between them. In essence, this modification will enable *ARGweaver* to fully subsume *G-PhoCS* as a coalescent-based inference framework that allows for both recombination and a general demographic model. Taking advantage of these new methods, we are working with various collaborators to identify older archaic introgression events. We anticipate that these studies will reveal important events in the evolution of human populations and lead to improved demographic models for use in studies of disease association and natural selection.

In addition, we have been working to make the powerful but complex *ARGweaver* program more accessible to users of software for population genetic analysis. Toward this end, Melissa recently wrote a detailed user's guide, with an accompanying github repository, that will be published as part of a book on statistical population genetics (Hubisz and Siepel, in press).

Analysis of Natural Selection on Regulatory Sequences in the Human Genome

We have a long-standing interest in characterizing the influence of natural selection on DNA sequences, particularly in noncoding regions of the genome. Most of our work in this area has involved comparisons of complete mammalian genomes and, hence, has considered evolutionary processes spanning 10s to 100s of millions of years. More recently, however, we have become interested in integrating this phylogenetic information with data on human polymorphism to gain insight into more recent evolutionary events. A few years ago, we developed a probabilistic model and inference method, called *INSIGHT*, that makes use of joint patterns of divergence and polymorphism to shed light on recent natural selection. We have used *INSIGHT* to show that natural selection has profoundly influenced transcription factor binding sites across the genome during the past five million years of evolution, with major contributions both to adaptive changes in humans and weakly deleterious variants currently segregating in human populations.

INSIGHT provides an estimate of the fraction of nucleotides under natural selection in any given collection of genomic elements. This same estimate can alternatively be interpreted as probabilities that mutations falling in the given elements will have fitness consequences for the organisms that carry them. We recently realized that this property could be used to

produce "fitness consequences" (*fitCons*) scores across the entire human genome. Using high-throughput data from the ENCODE project, we first partition the genome into classes of sites having characteristic functional genomic "fingerprints" in a given cell type. We then use *INSIGHT* to calculate a *fitCons* score for each fingerprint (Gulko et al., *Nat Genet 47:* 276 [2015]). Finally, we plot these scores along the genome sequence. These *fitCons* scores turn out to be remarkably powerful for identifying unannotated regulatory elements in the human genome. We have also used *fitCons* scores to estimate that only ~7% of nucleotides in the human genome directly influence fitness.

In our latest published effort in this area, we have addressed a major shortcoming of *fitCons*, namely, that it does not scale up for use with large numbers of functional genomic covariates. We developed an alternative approach that bypasses the need for clustering genomic sites and, instead, assumes a linear-logistic relationship between covariates along the genome and parameters of the *INSIGHT* model (Huang et al. 2017). This method, called *LINSIGHT*, is extremely fast and scalable, enabling it to exploit the "big data" available in modern genomics. We have shown that *LINSIGHT* outperforms the best available methods in identifying human noncoding variants associated with inherited diseases (including *fitCons*). In addition, we have applied *LINSIGHT* to an atlas of human enhancers and shown that the fitness consequences at enhancers depend on cell type, tissue specificity, and constraints at associated promoters.

In parallel, Brad Gulko has recently devised a powerful alternative approach to the *fitCons* clustering problem that both scales well and avoids the linearity assumptions of *LINSIGHT*. His new algorithm, called *fitCons2*, builds a decision tree by repeatedly splitting classes of genomic sites in a manner that is guaranteed to increase a global measure of the "information" associated with natural selection. This approach enables us to consider dozens of genomic features while allowing for complex combinations among them. Initial results suggest that it leads to substantially improved power and accuracy in estimating fitness consequences and identifying noncoding functional elements. Brad has applied *fitCons2* to all of the data from Roadmap Epigenomics, providing 115 cell type–specific *fitCons* maps. An interesting side benefit of this approach is that it estimates how many bits of information any given annotation yields about genomic function, as measured by natural selection. Consequently, it can be used to rank potential high-throughput experiments by their informativeness. We are using this property to build a *fitCons* map for rice based on newly generated data from the Purugganan laboratory at NYU.

In addition, Yi-fei Huang in the group has begun to develop deep-learning methods that both allow for arbitrarily complex relationships among genomic features and make use of population genetic theory and the full-site frequency spectrum to estimate allele-specific selection coefficients at every nucleotide in the human genome. This approach unifies methods for deleterious variant prediction with methods for inferring distributions of fitness. We plan to use the allele- and nucleotide-specific selection coefficients estimated by Yi-fei's new "DeepINSIGHT" method to perform the first rigorous analysis of the distribution of fitness effects (DFEs) across the noncoding genome, considering various subsets of interest (e.g., enhancers vs. promoters, or lineage-specific vs. deeply conserved enhancers).

Characterization of Transcriptional Regulation Using GRO-seq

For several years, we have been working closely with John Lis's group on methods for interpreting data generated using their powerful GRO-seq (global run-on and sequencing) technology, which maps the positions of engaged RNA polymerases across the genome. It has gradually become clear that an unanticipated benefit of GRO-seq and improved version of the protocol, called PRO-seq, is that they are uniquely well suited for detecting so-called enhancer RNAs (or eRNAs) and, consequently, for identifying active enhancers and other regulatory elements in mammalian cells.

We took advantage of this property in a recent study of the transcriptional response in human cells to celastrol, a compound derived from traditional Chinese medicine that has potent anti-inflammatory, tumor-inhibitory, and obesity-controlling effects (Dukler et al. 2017). Our analysis of PRO-seq data for K562 cells revealed dramatic transcriptional effects soon after celastrol treatment at a broad collection of both coding and noncoding transcription units. We found that transcriptional activity was generally repressed by celastrol, but one distinct group of genes, enriched for

roles in the heat-shock response, displayed strong activation. In addition, using a regression approach, we were able to identify key transcription factors that appear to drive these transcriptional responses, including members of the E2F and RFX families. This study showed that a careful analysis of PRO-seq time-course data can disentangle key aspects of a complex transcriptional response, and it provided new insights into the activity of a powerful pharmacological agent.

In our latest work in this area (Danko et al. 2018), we made use of PRO-seq to perform the first comparative study of nascent transcription in primates. We mapped actively transcribing RNA polymerases in resting and activated $CD4^+$ T-cells in multiple human, chimpanzee, and rhesus macaque individuals, with rodents as outgroups. This approach allowed us to directly measure active transcription separately from posttranscriptional processes. We observed general conservation in coding and noncoding transcription, punctuated by numerous differences between species, particularly at distal enhancers and noncoding RNAs. Among other findings, we observed that transcription factor binding sites are a primary determinant of transcriptional differences between species, rates of evolutionary change are strongly correlated with long-range chromatin interactions, and adaptive nucleotide substitutions are associated with lineage-specific transcription. We also found that genes regulated by larger numbers of enhancers are more frequently transcribed at evolutionarily stable levels, despite reduced conservation at individual enhancers. At one locus, SGPP2, we were able to experimentally validate that multiple substitutions contribute to human-specific transcription by altering NF-κB binding sites in the human genome. Collectively, our findings suggest a pervasive role for evolutionary compensation across ensembles of enhancers that jointly regulate target genes.

PUBLICATIONS

Dukler N, Booth GT, Huang Y-F, Tippens N, Waters CT, Danko CG, Lis JT, Siepel A. 2017. Nascent RNA sequencing reveals a dynamic global transcriptional response at genes and enhancers to the natural medicinal compound celastrol. *Genome Res.* **27:** 1–14.

Huang Y-F, Gulko B, Siepel A. 2017. Fast, scalable prediction of deleterious noncoding variants from functional and population genomic data. *Nat Genet* **49:** 618–624.

Kondo S, Vedanayagam J, Mohammed J, Eizadshenass S, Pang N, Kan L, Aradhya R, Siepel A, Steinhauer J, Lai EC. 2017. New genes often acquire male-specific functions but rarely become essential in *Drosophila*. *Genes Dev* **31:** 1841–1846.

Mohammed J, Flynt AS, Panzarino AM, Mondal MMH, Siepel A, Lai EC. 2017. Deep experimental profiling of microRNA diversity, deployment, and evolution across the *Drosophila* genus. *Genome Res* **28:** 52–65.

In Press

Danko CG, Choate LA, Marks BA, Rice EJ, Zhong W, Chu T, Martins AL, Dukler N, Coonrod SA, Wojno EDT, et al. 2018. Dynamic evolution of regulatory element ensembles in primate $CD4^+$ T-cells. *Nat Ecol Evol* **2:** 537–548.

Fang H, Huang Y-F, Radhakrishnan A, Siepel A, Lyon GJ, Schatz MC. 2018. Scikit-ribo enables accurate estimation and robust modeling of translation dynamics at codon resolution. *Cell Syst* **6:** 180–191.e4.

Hubisz MJ, Siepel AC. 2018. ARGweaver. In *Statistical population genetics* (ed. Dutheil J). Springer, New York (in press).

COLD SPRING HARBOR LABORATORY FELLOWS

In 1986, Cold Spring Harbor Laboratory began a Fellows Program to encourage independent research by outstanding young scientists who, during their graduate studies, displayed exceptional promise of becoming leading scientists of the future. The purpose of this program is to provide an opportunity for these young scientists to work independently at the Laboratory for a period of 3 years on projects of their own choosing. Fellows are provided with a salary, research support, and technical assistance so that they can accomplish their goals free of distraction. The interactions among research groups at the Laboratory and the program of courses and meetings on diverse topics in biology contribute to a research environment that is ideal for scientific innovation by these Fellows. The CSHL Fellows Program has been tremendously successful and has served as a paradigm for several analogous programs at other institutions, most recently a Fellows Program sponsored by the National Institutes of Health.

The success of the program is apparent from the list of distinguished alumni. Most notably, Carol Greider—recipient of the 2009 Nobel Prize in Physiology or Medicine for her work on telomerase and telomere function—joined the Fellows Program in 1998. After completing her fellowship, Carol was on the CSHL faculty for 9 years, and she is currently the Daniel Nathans Professor and Director of Molecular Biology and Genetics at Johns Hopkins University School of Medicine.

The first CSHL Fellow, Adrian Krainer (1986), is currently a Professor at the Laboratory, and Chris Vakoc (2008) and Florin Albeanu (2008) are currently holding Associate Professor positions at CSHL. Scott Lowe (1995) is a Howard Hughes Medical Institute (HHMI) Investigator. After nearly 15 years on the CSHL faculty, he recently took on a Professorship at Memorial Sloan Kettering Cancer Center in New York City. Marja Timmermans (1998) was a member of the CSHL faculty for more than 17 years and recently accepted the Humboldt Professorship at the University of Tübingen. Eric Richards (1989) currently is the Vice President of Research and Senior Scientist at the Boyce Thompson Institute for Plant Research at Cornell University; David Barford (1991) is a Fellow of the Royal Society and Professor of Molecular Biology at the Institute of Cancer Research in London; Ueli Grossniklaus (1994) is Professor at the Institute of Plant Biology, University of Zürich, Switzerland; and Térence Strick (2000) left at the end of his fellowship to become a Group Leader at the Institut Jacques Monod in Paris. Lee Henry (2000) joined HHMI's Janelia Farm in Ashburn and joined a project headed by Thomas Südhof. Ira Hall (2004) is an Associate Professor and Associate Director of the Genome Institute at Washington University. Patrick Paddison, who had joined the Fellows Program in 2004, currently is an Associate Member at the Fred Hutchinson Cancer Research Center in Seattle, Washington.

Lingbo Zhang has been a Fellow at the Laboratory since 2013. He joined us from Harvey Lodish's laboratory at the Whitehead Institute of the Massachusetts Institute of Technology, where he studied the regulation of erythroid progenitor cell self-renewal. As a CSHL Fellow, Lingbo is conducting genetic and small-molecule screens to discover novel regulators of normal and aberrant stem cell biology. **Jason Sheltzer** has been a CSHL Fellow since 2015 after completing his graduate work in Angelika Amon's laboratory at MIT. His research focuses on studies of aneuploidy and how it impacts cancer progression.

GENETIC APPROACHES TO INVESTIGATE CANCER

J. Sheltzer P. Baruah A. Lin C. Scaduto
C. Giuliano D. Lukow J. Smith
K. John A. Palladino A. Vasudevan

The Sheltzer lab applies a variety of techniques, including chromosome engineering, CRISPR mutagenesis, and single-cell analysis, to address fundamental questions in cancer biology. We are particularly interested in exploring the role of gene dosage imbalances in cancer development and progression. Additionally, we apply bioinformatic approaches to investigate genomic alterations that are associated with cancer patient outcomes to identify better biomarkers for treatment and prognosis.

Discovery and Characterization of Genes Affecting Patient Survival in Cancer

J. Sheltzer, J. Smith

In human patients, certain tumors can be cured by surgery alone. Other tumors derived from the same tissue and classified at the same pathological stage will inevitably recur after surgery and often prove to be fatal. For instance, stage II colorectal cancer is typically treated by surgical resection with curative intent. However, ~20% of these tumors will recur following surgery, and patients with recurrent disease have a 5-year survival rate of only 30%. Although adjuvant chemotherapy can lower the risk of colorectal cancer relapse, the difficulty in identifying patients who would benefit from additional treatment, coupled with its debilitating side effects, has limited its use. In general, many clinical decisions are constrained by the poor understanding of the molecular features that differentiate potentially fatal and nonfatal human tumors. A greater understanding of the genes and biological pathways that drive tumor aggressiveness would improve patient risk stratification and allow for more accurate treatment decisions.

To further our understanding of the molecular alterations that underlie deadly malignancies, we analyzed the genomic profiles of 35,946 solid tumors from patients with known outcomes. Surprisingly, we found that mutations in common cancer driver genes are almost never associated with patient survival time. However, copy number changes in these same genes are broadly prognostic in multiple cancer types. Analysis of methylation, microRNA, mRNA, and protein expression in primary tumors defined several additional prognostic patterns, including signatures of tumor mitotic activity and tissue dedifferentiation, that remain strongly associated with patient outcome after correcting for common clinical parameters. In total, we believe that this analysis establishes a comprehensive resource for biomarker identification and also underscores the importance of gene copy number profiling in assessing clinical risk.

Genetic Investigation of Cancer Drug Targets

C. Giuliano, K. John, A. Lin, A. Palladino

Our bioinformatic analysis of cancer patient data identified hundreds of genes whose expression correlated with clinical prognosis. We therefore set out to explore whether any of these genes could serve as a suitable target for drug development. To test this, we applied CRISPR-Cas9-based gene editing to test whether mutations in a particular gene would cause a loss in cell fitness. As a positive control in these assays, we included the maternal embryonic leucine zipper kinase (MELK), a gene that we identified as strongly associated with patient prognosis and had previously been implicated as a potent cancer drug target. Our assays involving MELK yielded very unexpected results, leading us to shift our attention to characterizing this gene in more detail.

MELK had been reported to be a genetic dependency in several cancer types, as small-molecule inhibitors of MELK and RNA interference (RNAi) against MELK block cancer cell proliferation. On the basis of those preclinical results, a MELK inhibitor is currently undergoing clinical trials in multiple cancer types. However, we found that mutagenizing MELK with CRISPR-Cas9 had no effect on the fitness of

cell lines from seven different cancer types. Cells that harbored null mutations in MELK showed wild-type doubling times, cell cycle progression, and anchorage-independent growth. Furthermore, MELK knockout lines remained sensitive to the MELK inhibitor in clinical trials, suggesting that this drug blocks cell division through an off-target mechanism. In total, these results undermined the rationale for a series of current clinical trials and provided an experimental approach for the use of CRISPR-Cas9 in preclinical target validation that can be broadly applied.

Because of the strong correlation between MELK expression and patient outcome, we continued to search for a function for MELK in cancer. We generated additional knockout clones of MELK, but found that, across cancer types, cells lacking MELK show wild-type growth in vitro, under environmental stress, in the presence of cytotoxic chemotherapies, and in vivo. By combining our MELK-knockout clones with a recently described, highly specific MELK inhibitor, we further showed that the acute inhibition of MELK results in no specific antiproliferative phenotype. Finally, we were able to discover why our work had identified MELK in the first place: A careful analysis of cancer gene expression showed that MELK expression was a correlate of tumor mitotic activity, explaining its association with poor clinical prognosis. Thus, tumors expressing high levels of MELK had a large fraction of cells in M phase, although the expression of MELK had no causative role in mitotic progression. Thus, we believe that these results conclusively explain why MELK is found in deadly tumors, and also show that MELK is absolutely nonessential for cancer.

Chromosomal Instability and Aneuploidy in Cancer

P. Baruah, D. Lukow, C. Scaduto, J. Sheltzer, A. Vasudevan

More than 90% of human tumors display whole-chromosome copy number changes, or aneuploidies, that alter the dosage of hundreds or thousands of genes at once. Despite the ubiquity of aneuploidy in cancer, we have little understanding of how these dosage imbalances affect tumor development. We are therefore applying multiple genetic approaches to model this condition and shed light on the role that it plays in cancer.

To test the link between aneuploidy and cancer, we transduced congenic euploid and trisomic fibroblasts with 14 different oncogenes or oncogene combinations, thereby creating matched cancer cell lines that differ only in karyotype. Surprisingly, nearly all aneuploid cell lines divided slowly in vitro, formed few colonies in soft agar, and grew poorly as xenografts, relative to matched euploid lines. Similar results were obtained when comparing a near-diploid human colorectal cancer cell line with derivatives of that line that harbored extra chromosomes. Only a few aneuploid lines grew at close to wild-type levels, and no aneuploid line showed greater tumorigenic capabilities than its euploid counterpart. These results show that rather than promoting tumorigenesis, aneuploidy can very often function as a tumor suppressor. Moreover, our results suggest one potential way that cancers can overcome the tumor suppressive effects of aneuploidy: On prolonged culture in vitro or in vivo, cell lines with simple aneuploidies developed recurrent chromosomal aberrations that were absent from their euploid counterparts and were associated with enhanced growth. Thus, the genome-destabilizing effects of single-chromosome aneuploidy may facilitate the evolution of balanced, high-complexity karyotypes that are frequently found in advanced malignancies.

These results raised a central question: If aneuploidy is detrimental to cancer cell proliferation, then why are tumors aneuploid? We are investigating several different possibilities to explain this phenomenon. First, we are seeking to uncover whether a link exists between aneuploidy and metastasis. Using our colon cancer cell lines harboring extra chromosomes, we found that adding a single extra copy of chromosome 5 showed a partial epithelial–mesenchymal transition. These cells strongly down-regulated E-cadherin, Epcam, and Claudin-7 and showed increased motility and invasive behavior. These patterns were not observed in cells harboring several additional aneuploidies, suggesting that the phenotype is caused by the increased dosage of a gene or genes found on chromosome 5. We are applying a variety of genetic approaches to identify these key factors.

More generally, we hypothesize that aneuploidy is commonly detrimental under "normal" growth conditions. That is, when a cell is grown in rich media with an adequate supply of nutrients and growth factors, aneuploidy is disfavored. However, in stressful environments, unique karyotypes may exist that confer an environment-specific growth advantage. To test this,

we have treated cancer cells with Mps1 inhibitors to generate populations of cells with random aneuploidies. We have observed that pretreatment with Mps1 inhibitors speeds the evolution of drug resistance in cells exposed to various chemotherapy agents. In the case of one drug, vemurafenib, this resistance consistently co-occurs along with the gain of chromosomes 11 and 18. We speculate that similar aneuploidy patterns may exist for other drugs or environments—that is, an "optimal" karyotype can be found to maximize growth potential in each condition. We believe that these results may explain the close association that we have previously documented between aneuploidy and poor prognosis in cancer.

PUBLICATIONS

Lin A, Giuliano CJ, Sayles NM, Sheltzer JM. 2017. CRISPR/Cas9 mutagenesis invalidates a putative cancer dependency targeted in on-going clinical trials. *eLife* **6:** e24179.

Sheltzer JM, Ko JC, Replogle JM, Habibe Burgos NC, Chung ES, Meehl CM, Sayles NM, Passerini V, Storchova Z, Amon A. 2017. Single-chromosome gains commonly function as tumor suppressors. *Cancer Cell* **31:** 1–16.

Smith JC, Sheltzer JM. 2017. An integrated analysis of the epigenetic, genetic, and transcriptional patterns associated with outcome across cancer types. bioRxiv doi: 10.1101/186528.

In Press

Giuliano CJ, Lin Al, Smith JC, Palladino AC, Sheltzer JM. 2018. MELK expression correlates with tumor mitotic activity but is not required for cancer growth. *eLife* **7:** e32838.

TARGETING SELF-RENEWAL PATHWAYS AND METABOLIC DEPENDENCIES IN HEMATOPOIETIC DISEASES AND MALIGNANCIES

L. Zhang J. Korossy W. Mei D. Tsang Hu E. Wong
B. Li G. Trivedi K. Vishnu L. Zhu

The research in our laboratory addresses a central question in blood cell formation—a process also known as hematopoiesis—which is how self-renewal and differentiation are properly balanced in the hematopoietic stem and progenitor cell (HSPC) population. The balance is crucial because dysregulation of this process causes various hematopoietic diseases and malignancies. Our laboratory uses both CRISPR-Cas9 functional genomic and forward chemical genomic approaches to identify novel regulators of HSPC self-renewal, and we aim to develop novel therapeutic strategies for hematopoietic diseases and malignancies, including myelodysplastic syndrome (MDS) and acute myeloid leukemia (AML).

Thus far, we have identified a G protein–coupled receptor (GPCR) as a novel druggable target to regulate self-renewal of early erythroid progenitor. We have shown that pharmacological inhibition of this GPCR completely corrected MDS in a genetically engineered *Mx1-Cre Srsf2P95H/WT* MDS-diseased mouse model that faithfully recapitulates essential pathological features of human MDS. Through collaboration with the Northwell Health System, we are currently developing highly selective antagonists of this GPCR as lead compounds for clinical trials to treat MDS. In another project focusing on leukemic cells, we have identified metabolic enzyme pyridoxal kinase (PDXK) as an AML selective dependency. We have shown that inhibition of PDXK specifically disrupted leukemia maintenance in vivo while maintaining minimal side effects toward normal HSPCs. As PDXK catalyzes the formation of pyridoxal phosphate (PLP), which is the bioactive form of vitamin B_6, our work also uncovered an unexpected addiction of AML on vitamin B_6. Through collaborations with medicinal chemists, we are developing small chemical compound inhibitors to better drug PDXK as novel therapeutics for AML.

Targeting Early Erythroid Progenitor Deficiency in Refractory MDS

MDS is a form of lethal hematopoietic malignancy that is characterized by pancytopenia resulting from progressive bone marrow failure. Clinically, the therapeutic options for MDS are very limited. Only ~20% of MDS patients benefit from standard erythropoietin (EPO) treatment, and many of the initial responders do not have a long-term response. One major advance in MDS treatment in recent years is the approval of lenalidomide as a novel drug to specifically treat a subgroup of MDS with chromosome 5q deletion, which only accounts for ~5% of the total MDS population. The only option for patients who do not respond to EPO and lenalidomide is red blood cell (RBC) transfusion, but transfusion exposes patients to insufficient correction of anemia, alloimmunization, and organ failure secondary to iron overload. Therefore, there is an unmet and urgent demand for patients and clinicians to have novel therapeutics to treat these refractory MDSs.

We discovered a GPCR as novel regulator of early erythroid progenitor self-renewal and showed that pharmacological inhibition of this critical receptor completely corrected bone marrow failure of MDS in genetically engineered MDS mouse models and cultured primary human patient samples of MDS. This receptor is expressed in the burst-forming unit erythroid (BFU-E) progenitor and is up-regulated during erythroid differentiation and down-regulated during self-renewal. In the primary BFU-E progenitor culture system, genetic down-regulation of this receptor or pharmacologic inhibition of it using nanomolar concentrations of its selective antagonists promoted BFU-E progenitor self-renewal. Strikingly, injection of its antagonists completely corrected anemias of MDS in vivo in the genetically accurate MDS mouse model *Mx1-Cre Srsf2P95H/WT*, which faithfully recapitulates the essential pathological phenotypes of MDS. The treatment also extended survival of MDS mice to the life

span of control wild-type mice. Mechanistically, we found that cAMP response element-binding protein (CREB), a downstream effector of the receptor, preferentially binds to and triggers the up-regulation of genes important for the maintenance of BFU-E progenitor status. These genes include the transcription factor GATA2, haploinsufficiency of which causes pediatric MDS, and ZFP36L2, an RNA-binding protein that we recently identified as a crucial molecular switch, balancing BFU-E self-renewal and differentiation.

We further demonstrated that similar to human MDS patients, MDS mouse models showed abnormal reduction of BFU-Es and elevation of plasma EPO levels in comparison to wild-type control mice. Importantly, injection of receptor inhibitors corrected both BFU-E deficiency and the abnormal plasma EPO levels of MDS mice to levels comparable to wild-type control mice. These results showed that targeting this receptor overcame BFU-E deficiency and EPO refractoriness. We have received multiple awards for these promising preclinical advances, including the Evans Foundation MDS Young Investigator Award and the National Institutes of Health (NIH) Research Evaluation and Commercialization Hub (REACH) Awards, which include the phase I Feasibility Award and phase II Proof-of-Concept Award.

Our laboratory is currently translating these promising preclinical discoveries into clinical trials for EPO-refractory MDS. Having shown that the selective antagonists corrected anemia of MDS in the Srsf2P95H genetic background, we will further assess the efficacy of the inhibition in multiple murine genetic models of MDS carrying different genetic abnormalities and identify genetic markers of response to the antagonists using primary MDS patient samples. To maximize therapeutic efficacies and reduce nonspecific side effects of these antagonists, we will modify the structure of these antagonists to optimize drug metabolism and pharmacokinetic (DMPK) properties. Through collaboration with the Northwell Health System, we are currently developing highly selective antagonists of this GPCR as lead compounds for clinical trials to treat MDS.

Targeting Metabolic Abnormality as Novel Cancer Dependencies in Acute Myeloid Leukemia

Cancer cells undergo metabolic reprogramming to support their abnormal proliferation. However, because of our limited understanding of molecular details and lack of validated drug targets, designing therapeutic strategies to exploit aberrant metabolism has proven challenging. The recent success of development of isocitrate dehydrogenase (IDH) inhibitors as novel agents to treat AML highlights the potential of treating AML through targeting dysregulated metabolism in cancer cells. However, the vast majority of oncogenic drivers alter cellular metabolism through indirect mechanisms, and, consequently, the metabolic regulators that causally contribute to cancer initiation and maintenance are not obvious and cannot be inferred from genomic analyses alone. More importantly, normal stem and progenitor cells in adult tissues, such as HSPCs, also undergo regular expansion and proliferation, thereby sharing similar metabolic requirements as cancer cells, such as the extensive formation of biomacromolecules to generate building blocks for daughter cell formation. Consequently, one major side effect of most standard chemotherapies is the depletion of HSPCs, which resulted from targeting shared metabolic features. Therefore, design of novel therapeutics selectively targeting leukemic cells is challenging.

Through analyzing gene expression profiles from $CD34^+$ HSPCs and leukemic cells, we found that the global metabolic gene expression signature is capable of distinguishing these two cell types. Importantly, the metabolic difference among different subtypes of AML cells carrying different genetic abnormalities is significantly smaller than the difference between them and HSPCs. This suggests that there are common metabolic vulnerabilities shared by multiple subtypes of acute leukemia, which potentially serve as drug targets to block their proliferation, while showing no or minimal side effect on HSPCs. We identified approximately 300 metabolic genes that follow this pattern in which they are highly expressed across multiple subtypes of leukemia in comparison to HSPCs. To identify metabolic vulnerabilities in AML not harboring IDH mutations, we performed a focused CRISPR-Cas9 "drop-out" screen using a single guide RNA (sgRNA) library targeting metabolic genes highly expressed in AML cells.

We identified PDXK—an enzyme that produces PLP from vitamin B6—to be preferentially required for AML cell proliferation compared with many other cell types. PDXK kinase activity is required for both PLP production and proliferation of AML cells, and

pharmacological suppression of PLP with the antituberculosis drug isoniazid recapitulated the effects of PDXK disruption. PLP is a cofactor for many enzymes involved in cell proliferation and, accordingly, PDXK disruption reduced intracellular concentrations of key metabolites needed for cell division, and disruption of the PLP-dependent enzymes ornithine decarboxylase (ODC1) or glutamic-oxaloacetic transaminase 2 (GOT2) selectively inhibited AML cell proliferation in a manner that could be partially rescued by addition of downstream products.

Our work identified the vitamin B_6 pathway as a pharmacologically actionable dependency in AML. We revealed that leukemia cells are addicted to the vitamin B_6 pathway such that its inhibition selectively impairs their proliferation compared with other normal and cancer cell types. Our results support the emerging view that essential vitamins (e.g., vitamin D in pancreatic cancer or vitamin C in colorectal cancer and leukemia) can ultimately play decisive roles in cancer cell proliferation and maintenance and align with recent epidemiological studies hinting that vitamin B_6 is not chemopreventive but instead may increase cancer risk. Notable is that ODC1 has previously been identified as a target in colon and several other cancers, and that one biochemical outcome of GOT2 inhibition is a reduction in asparagine, which can also be achieved in leukemia by the U.S. Food and Drug Administration (FDA)-approved drug asparaginase. These observations imply that the vitamin B_6 pathway coordinates multiple activities that are critical for cancer maintenance, and PDXK inhibitors, by simultaneously attenuating these pathways, are important antileukemia agents.

Functional Genomics Screen to Identify Regulatory Network of Early Erythroid Progenitor Self-Renewal

Erythropoiesis is a multistage developmental process that results in erythrocyte production. EPO is a crucial regulator of this process, triggering survival of the late erythroid progenitor, the colony-forming unit erythroid (CFU-E), and has thus been widely used for the treatment of anemias. However, many anemias associated with cancer chemotherapy and radiation therapy are refractory to EPO treatment. The refractoriness of these pathologic conditions is partially caused by insufficiency and loss of the BFU-E, an early erythroid progenitor that is unresponsive to EPO. Therefore, targeting BFU-E self-renewal is crucial for sustained erythropoiesis and treatment of anemias in these conditions. A systematic and functional analysis to better understand molecular mechanisms underlying BFU-E self-renewal holds promise in the development of novel therapeutics for EPO-refractory anemias associated with cancer chemotherapy and radiation therapy.

It has been shown that gene transactivation of the glucocorticoid receptor (GR) is essential for erythroid lineage expansion, and we have previously identified RNA-binding protein ZFP36L2 as a direct GR target gene essential for this process. ZFP36L2 binds to messenger RNAs (mRNAs) highly expressed during erythroid differentiation and negatively regulates their expression. However, although several essential GR downstream target genes and cofactors have been identified, a systematic understanding of how the GR regulates BFU-E self-renewal is still missing. Here, we used glucocorticoid-induced gene expression profiling, GR chromatin immunoprecipitation sequencing (ChIP-seq), and short hairpin RNA (shRNA) functional genomic screening to provide a systematic and functional understanding of molecular mechanisms underlying BFU-E self-renewal. We combined glucocorticoid-induced differential gene expression profiling, GR ChIP-seq, and shRNA functional genomic screening to identify 43 genes as direct targets of the GR and 21 genes as novel regulators of BFU-E self-renewal. We further validated several identified genes as negative regulators of BFU-E self-renewal. Our results not only provide a systematic and functional understanding of the BFU-E self-renewal mechanism, but they also serve as a resource for future research into novel BFU-E self-renewal regulators.

In summary, our research centers on self-renewal mechanisms, and we investigate how self-renewal is regulated and how dysregulation of this process causes diseases. Through both genetic and chemical functional genomics, our research identified a GPCR and the metabolic enzyme PDXK as novel druggable regulators of self-renewal of early erythroid progenitor and leukemia cells, respectively. We are currently collaborating with medicinal chemists to further modify DMPK properties of small chemical inhibitors to

better target self-renewal in pathological conditions of MDS and AML. Through collaboration with hospitals and clinicians, we aim to test our lead compounds in primary disease human samples of MDS and AML and plan to move the most promising lead compounds into clinical trials to treat these two lethal diseases.

PUBLICATIONS

Lowe S, Zhang L, Chen C. 2017. Compositions and methods for treating acute myeloid leukemia. Patent filed.

Zhang L, Trivedi G. 2017. Compositions and methods for treating anemia. Patent filed.

AUTHOR INDEX

WATSON SCHOOL OF BIOLOGICAL SCIENCES

ADMINISTRATION

Alexander Gann, Ph.D., WSBS Professor and Dean
Monn Monn Myat, Ph.D., Associate Dean
Alyson Kass-Eisler, Ph.D., Postdoctoral Program Officer and Curriculum Director
Kimberly Creteur, M.Ed., M.S.Ed., Admissions and Recruitment Manager
Kimberley Geer, Administrative Assistant

EXECUTIVE COMMITTEE

Chair
Alexander Gann

Members
Mikala Egeblad
Alexander Krasnitz
Bo Li
David L. Spector (Director of Research)
David Stewart
Anthony Zador

Student Representatives
Jacqueline Giovanniello, WSBS
Dongyan Song, SBU

Secretary
Alyson Kass-Eisler
Monn Monn Myat

ADMISSIONS COMMITTEE

Chair
Adrian Krainer

Members
Alexander Gann (ex officio)
Bo Li
Zachary Lippman
W. Richard McCombie
Stephen Shea
Adam Siepel
Nicholas Tonks
Christopher Vakoc
Linda Van Aelst

Secretary
Kimberly Creteur

QUALIFYING EXAM COMMITTEE

Chair
Linda Van Aelst

Members
Molly Hammell
W. Richard McCombie
Stephen Shea
Christopher Vakoc

EXTERNAL ADVISORY COMMITTEE

Keith Yamamoto (Chair)
Executive Vice Dean, School of Medicine
University of California, San Francisco

Victor Corces
Professor, Department of Biology
Emory University
Professor, Howard Hughes Medical Institute

Gail Mandel
Senior Scientist, Vollum Institute
Oregon Health and Science University
Investigator, Howard Hughes Medical Institute

Marguerite Mangin
Academic Programs Director and Senior Research Associate
The Rockefeller University

Barbara Meyer
Professor of Genetics and Development
University of California, Berkeley
Investigator, Howard Hughes Medical Institute

Frank Solomon
Professor, Department of Biology and Center for Cancer Research
Massachusetts Institute of Technology

WATSON SCHOOL OF BIOLOGICAL SCIENCES DEAN'S REPORT

The 13th WSBS Graduation

On May 7, 2017, we celebrated the Watson School of Biological Sciences (WSBS) 13th graduation ceremony. Eight students were awarded Ph.D. degrees: Matt Koh from the Entering Class of 2010; Brittany Cazakoff, Joaquina Delas Vives, Justus Kebschull, and Charles Underwood from the Entering Class of 2011; Annabel Romero Hernandez and Abram Santana from the Entering Class of 2012; and Maria Nattestad from the Entering Class of 2013. An Honorary degree was bestowed upon Dr. Carol Greider.

Carol earned her Ph.D. at the University of California, Berkeley, where, as a graduate student working with Dr. Elizabeth Blackburn, she discovered telomerase, an enzyme that maintains telomeres, the "caps" at the ends of chromosomes. Carol came to Cold Spring Harbor Laboratory in 1988, where, as the Lab's second CSHL Fellow, she cloned and characterized the RNA component of telomerase. In 1997, Carol moved her laboratory to the Johns Hopkins University School of Medicine. She was awarded the 2009 Nobel Prize for Physiology or Medicine (shared with Dr. Blackburn and Dr. Jack Szostak). She is currently the Daniel Nathans Professor and Director of Molecular Biology and Genetics at Hopkins. In her convocation address, Carol's chief message to the graduates was "whenever I get knocked down, I get back up again." She advised the graduates to expect both ups and downs in the research careers that lay before them—but never to become discouraged.

Dr. Winship Herr, the founding Dean of the Watson School, also participated in the ceremony, providing brief words of encouragement to the graduating class and citing the achievements of the School's graduates as a clear mark of its success. Dr. Herr is planning on writing a book on the founding of the School upon his retirement from the University of Lausanne next year.

2017 Graduates: (*Left* to *right*) CSHL President Bruce Stillman, honorary degree recipient Carol Greider, Maria Nattestad, Justus Kebschull, CSHL Chancellor Emeritus James Watson, Charles Underwood, Abram Santana, Annabel Romero Hernandez, Joaquina Delas Vives, Brittany Cazakoff, WSBS Dean Alex Gann.

2017 WSBS DOCTORAL RECIPIENTS

Student	Thesis advisor	Academic mentor	Current position
M. Joaquina Delás Vives	Gregory Hannon	Nicholas Tonks	Postdoctoral Fellow, Cambridge University (Advisor: Gregory Hannon)
Abram Handly-Santana	David Tuveson	Lloyd Trotman	Postdoctoral Fellow, Columbia University (Advisor: Christine Chio)
Yu-Jui (Ray) Ho	Molly Hammell	Michael Schatz	Computational Biologist, Memorial Sloan Kettering Cancer Center (Advisor: Scott Lowe)
Justus Kebschull	Anthony Zador	Bruce Stillman	Postdoctoral Fellow, Stanford University (Advisor: Liqun Luo)
Fred Marbach	Anthony Zador	Josh Dubnau	Postdoctoral Fellow, Sainsbury Wellcome Centre, United Kingdom (Advisor: Marcus Stephenson-Jones)
Maria Nattestad	Michael Schatz	Linda Van Aelst	Senior Computational Biologist, Fluidigm, California
Onyekachi Odoemene	Anne Churchland	Stephen Shea	Postdoctoral Fellow, Army Research Laboratory, Maryland (Advisor: Chou Hung)
Annabel Romero Hernandez	Hiro Furukawa	Adrian Krainer	Scientist, Regeneron, New York

2017 THESIS DISSERTATION DEFENSES

ENTERING CLASS OF 2011

M. Joaquina Delás Vives, March 24, 2017

Role of long noncoding RNAs in hematopoiesis.

Thesis Examining Committee

Chair:	**Christopher Vakoc**
Research Mentor:	**Gregory Hannon**
Academic Mentor:	**Nicholas Tonks**
Committee Member:	**Adrian Krainer**
External Examiner:	**John Rinn** *Harvard University*

Justus Kebschull, March 31, 2017

Exploiting DNA sequencing technology for high-throughput neuroanatomy.

Thesis Examining Committee

Chair:	**Bo Li**
Research Mentor:	**Anthony Zador**
Academic Mentor:	**Bruce Stillman**
Committee Member:	**Josh Dubnau**
External Examiner:	**Hongui Zeng** *Allen Institute for Brain Science*

Fred Marbach, July 13, 2017

Projection-neuron activity in auditory cortex during perceptual decisions.

Thesis Examining Committee

Chair:	**Alexei Koulakov**
Research Mentor:	**Anthony Zador**
Academic Mentor:	**Josh Dubnau**
Committee Member:	**Florin Albeanu**
External Examiner:	**Robert Froemke** *New York University*

Onyekachi Odoemene, June 29, 2017

Accumulation of visual evidence for decision-making in mice.

Thesis Examining Committee

Chair:	**Anthony Zador**
Research Mentor:	**Anne Churchland**
Academic Mentor:	**Stephen Shea**
Committee Member:	Z. Josh Huang
External Examiner:	**Lindsey Glickfeld** *Duke University*

ENTERING CLASS OF 2012

Abram Handly-Santana, April 21, 2017

Distinct populations of fibroblasts in pancreatic cancer.

Thesis Examining Committee

Chair:	**Christopher Vakoc**
Research Mentor:	**David Tuveson**
Academic Mentor:	**Lloyd Trotman**
Committee Member:	**Darryl Pappin**
External Examiner:	**Ellen Puré** *University of Pennsylvania*

(*continued*)

2017 THESIS DISSERTATION DEFENSES (*continued*)

Yu-Jui (Ray) Ho, July 12, 2017

Single-cell sequencing to understand resistance mechanisms in melanoma.

Thesis Examining Committee

Chair:	**David Tuveson**
Research Mentor:	**Molly Hammell**
Academic Mentor:	**Michael Schatz**
Committee Member:	**Mickey Atwal**
Committee Member:	**Michael Wigler**
External Examiner:	**Guillaume Bourque** *McGill University*

ENTERING CLASS OF 2013

Maria Nattestad, January 26, 2017

Computational methods for analysis and visualization of long-read sequencing data in cancer genomics.

Thesis Examining Committee

Chair:	**W. Richard McCombie**
Research Mentor:	**Michael Schatz**
Academic Mentor:	**Linda Van Aelst**
Committee Member:	**Adam Siepel**
External Examiner:	**Adam Hillippy** *National Human Genome Research Institute*

DOCTORAL THESIS RESEARCH

Student	Academic mentor	Research mentor	Thesis research
ENTERING CLASS OF 2012			
Talitha Forcier *NIH Predoctoral Trainee* *William Randolph Hearst Foundation Scholar*	Nicholas Tonks	Justin Kinney	Building the transcriptional regulatory code from the ground up.
Paul Masset *Florence Gould Fellow*	Jan A. Witkowski	Adam Kepecs	Representations of decision confidence in the brain: From Bayes' rule to channelrhodopsin.
ENTERING CLASS OF 2013			
Giorgia Battistoni *Starr Centennial Scholar* *Boehringer Ingelheim Fonds Fellow*	Christopher Hammell	Gregory Hannon	One-carbon metabolism and DNA methylome dynamics in pancreatic cancer.
Lital Charatifsky *Israeli WSBS Fellow*	John Inglis	Anne Churchland	Neural circuits for multisensory integration in normal and disease states.
Sanchari Ghosh *Charles A. Dana Fellow*	Josh Dubnau	Anthony Zador	Role of corticostriatal plasticity in learning an auditory discrimination task.
Michael Gutbrod *Bristol-Myers Squibb Fellow*	Zachary Lippman	Robert Martienssen	Small RNA and the RNAi pathway in transposable element regulation and differentiation in the preimplantation embryo.
Daniel Kepple *Crick-Clay Fellow* *NIH Predoctoral Trainee*	David Stewart	Alexei Koulakov	The human olfactory space.
Laura Maiorino *George A. and Marjorie H. Anderson Fellow* *Boehringer Ingelheim Fonds Fellow*	Nicholas Tonks	Mikala Egeblad	Understanding the role of the epithelial–mesenchymal plasticity in pancreatic cancer metastasis.
Georgi Yordanov *Leslie C. Quick, Jr. Fellow* *Boehringer Ingelheim Fonds Fellow*	Leemor Joshua-Tor	David Tuveson	Role of c-Myc in pancreatic cancer.
ENTERING CLASS OF 2014			
Emillis Bruzas *Starr Centennial Scholar*	Alea Mills	Mikala Egeblad	Investigation of mechanisms responsible for reawakening and chemoresistance in a breast cancer dormancy model.
Hamza Giaffar *Robert and Teresa Lindsay Fellow*	Jan Witkowski	Alexei Koulakov	The primacy model of olfactory coding.

(*continued*)

DOCTORAL THESIS RESEARCH (*continued*)

Student	Academic mentor	Research mentor	Thesis research
Jacqueline Giovanniello *NIH Predoctoral Trainee*	Bruce Stillman	Bo Li	Disruption of central amygdala fear circuit in a 16p11.2 microdeletion model of autism.
Elizabeth Hutton *Elizabeth Sloan Livingston Fellow*	Molly Hammell	Adam Siepel	Functional variant prediction in noncoding regions.
Sashank Pisupati *Cashin Fellow*	Stephen Shea	Anne Churchland	Dissecting the circuits and mechanisms that support optimal multisensory integration in rodents.
Colin Stoneking *NIH Predoctoral Trainee*	Zachary Lippman	Anthony Zador	Neuronal mechanisms enabling decision-making to be learned.
Jue Xiang Wang *George A. and Marjorie H. Anderson Fellow* *Boehringer Ingelheim Fonds Fellow*	Mikala Egeblad	Hiro Furukawa	Impact of subunit composition and de novo mutations on NMDA receptor structure, channel function, and interactions.
Anqi Zhang *Starr Centennial Scholar*	Bo Li	Anthony Zador	From corticostriatal plasticity to a common pathway.
ENTERING CLASS OF 2015			
Benjamin Berube *National Science Foundation Fellow* *NIH Predoctoral Trainee* *Elizabeth Sloan Livingston Fellow*	Zachary Lippman	Robert Martienssen	A single-cell assessment of germline epigenetic heterogeneity.
Kristina Grigaityte *Farish-Gerry Fellow*	John Inglis	Mickey Atwal	Computational analyses of high-throughput single-cell T-cell receptor sequences in health and disease.
Matt Lee *David H. Koch Fellow*	Nicholas Tonks	Lloyd Trotman	The road to metastasis: Defining the initial stages of prostate cancer progression.
Katarina Meze *Leslie C. Quick, Jr. Fellow*	Jay Lee	Leemor Joshua-Tor	Structural and functional studies of RNA regulatory mechanisms mediated by Lin28.
Alexandra Nowlan *Genentech Fellow* *George A. and Marjorie H. Anderson Fellow*	Jessica Tollkuhn	Stephen Shea	Multisensory experience-dependent plasticity: Network dynamics in auditory processing following parturition.
Sofya Polyanskaya *Starr Centennial Scholar*	Alexander Krasnitz	Christopher Vakoc	Identification of fusion oncoprotein codependencies in cancer.
Ngoc Tran *Samuel Freeman Fellow*	Leemor Joshua-Tor	Alexei Koulakov	Predicting bioactivity using a data-driven representation of three-dimensional chemical structures.
ENTERING CLASS OF 2016			
Brianna Bibel *National Science Foundation Fellow*	Hiro Furukawa	Leemor Joshua-Tor	Structural and functional studies of phosphorylation-mediated regulation of the RNAi effector Argonaute.
Alberto Corona *NIH Predoctoral Trainee* *Hearst Foundation Fellow*	David Jackson	Stephen Shea	Identification of neural circuitry underlying paternal behaviors.
David Johnson *National Science Foundation Fellow* *Hearst Foundation Scholar*	Zachary Lippman	Alea Mills	Elucidating the role of BRPF1 in human glioblastoma multiforme.
Christopher Krasniak *NIH Predoctoral Trainee*	Jan Witkowski	Anthony Zador	The role of cholinergic input to visual cortex in mouse spatial visual attention.
Shaina Lu *Edward and Martha Gerry Fellow*	Leemor Joshua-Tor	Anthony Zador	Development of a high-throughput pipeline to study the relationship of neuron projections and gene expression underlying mouse models of neuropsychiatric disorders.
Kathryn O'Neill *NIH Predoctoral Trainee*	Camila dos Santos	Ullas Pedmale	Cryptochrome-mediated responses to light signaling through RNA N6- methyladenosine modifications.

(*continued*)

DOCTORAL THESIS RESEARCH (*continued*)

Student	Academic mentor	Research mentor	Thesis research
Luqun Shen *Edward and Martha Gerry Fellow*	David Stewart	Stephen Shea	Predicting bioactivity using a data-driven representation of three-dimensional chemical structures.
Olya Spassibojko *David and Fanny Luke Fellow*	Jessica Tollkuhn	Ullas Pedmale	Molecular determinants controlling cryptochrome light signal transduction.
Martyna Sroka *George A. and Marjorie H. Anderson Fellow*	Molly Hammell	Christopher Vakoc	Molecular dissection of the PAX3-FOXO1 fusion oncoprotein pathway in rhabdomyosarcoma.
Ran Yan *George A. and Marjorie H. Anderson Fellow*	David Tuveson	Douglas Fearon	Identification of endogenous antigen-specific T cells in pancreatic cancer metastasis.
Chengxiang (Charlie) Yuan *A*STAR Fellow*	Nicholas Tonks	Jay Lee	Linking the cell cycle and developmental fate specification.

Teaching Award

Mickey Atwal

At this year's graduation ceremony, the School awarded its 11th Winship Herr Award for Excellence in Teaching to Dr. Mickey Atwal, the lead instructor of the Specialized Disciplines course in Quantitative Biology. Mickey, who also won the award in 2011 and 2016, was chosen by the first-year students for this award. Here is some of what the students said about Mickey in their nominations: "Mickey showed a strong ability to describe complicated concepts in intelligible ways to a broad audience. He also was proficient at targeting his pace and content to an appropriate level given his audience." And, "He helped explain complex topics in a variety of ways, including drawing different graphs, going through example problems with us, and challenging us with a hard but fun problem set."

Faculty and Administrative Changes

In August 2017, we welcomed Dr. Monn Monn Myat to the WSBS as our new Associate Dean. While an undergrad at Mount Holyoke College, Monn was an URP here in Adrian Krainer's lab in the summer of 1990. She went on to do her Ph.D. at Rockefeller with Alan Aderem and was a postdoc at Johns Hopkins with Deborah Andrew. She then took up a faculty position at Weill Cornell, where she was an Assistant and then Associate Professor for 10 years with an active research group and teaching responsibilities. Prior to joining the School, she spent 3 years as an Associate Professor at Medgar Evers College, part of the CUNY system, where she continued to teach and mentor undergraduates in her lab. Monn spent the fall semester auditing our first-year curriculum while also writing a grant to potentially launch a postbaccalaureate program at the WSBS.

Two new faculty members joined the Watson School in 2017: Tatiana Engel and David McCandlish. Tatiana earned her Ph.D. in physics at the Humboldt University of Berlin, followed by postdoctoral and research scientist positions in physics and computational neuroscience. Her expertise is in the analysis of high-dimensional neural data sets and in biophysical modeling. Her work at CSHL uses computational methods to investigate how perception and cognition arise from changes in neural activity, and she develops mathematical models to explain how these activity changes emerge from signaling between neurons, ultimately driving behavior.

David earned his Ph.D. in biology at Duke University, where he focused on understanding the structure of sequence-to-fitness mappings. As a postdoc, he studied the role of epistasis in protein evolution. His lab at CSHL develops computational tools to predict the functional impact of mutations in protein coding sequences.

Tatiana and David have already participated in WSBS activities, including giving Research Topics talks to the first-year students. David also served as an expert examiner on a Qualifying Exam. We look forward to their growing participation as members of the faculty.

Admissions 2017

The School received a very high-quality pool of more than 200 applications for the Entering Class of 2017 and is indebted to its Admissions Committee, which reviewed, interviewed, and selected candidates for our doctoral program. The Admissions Committee for the 2017 entering class comprised Adrian Krainer (Chair), Bo Li, Zachary Lippman, W. Richard McCombie, Stephen Shea, Adam Siepel, Nicholas Tonks, Christopher Vakoc, Linda Van Aelst, and myself (ex officio).

Entering Class of 2017

On August 21, 2017, the WSBS welcomed the 18th incoming class, consisting of eight new students: Lyndsey Aguirre, Sara Boyle, Bruno Gegenhuber, Benjamin Harris, Yuzhao (Richard) Hu, Dennis Maharjan, Diogo Maia e Silva, and Cole Wunderlich.

ENTERING CLASS OF 2017

Lyndsey Aguirre, The University of Texas, Austin: B.S. in Cell and Molecular Biology (2017)
Academic Mentor: Ullas Pedmale

Sara Boyle, Northwestern University: B.A. in Neuroscience (2017)
Academic Mentor: Jessica Tollkuhn

Jordan Gegenhuber, Pacific University: B.S. in Biology (2016)
Academic Mentor: John Inglis

Benjamin Harris, Colgate University: B.A. in Biology and Computer Science (2017)
Academic Mentor: W. Richard McCombie

Yuzhao "Richard" Hu, Tsinghua University: B.S. in Biological Sciences (2017)
Academic Mentor: Jay Lee

Dennis Maharjan, Brandeis University: M.S. in Molecular and Cellular Biology (2016); Caldwell University: B.S. in Biology (2012)
Academic Mentor: Florin Albeanu

Diogo Maia e Silva, University of Lisbon: M.D. in Medicine (2017)
Academic Mentor: Bruce Stillman

Cole Wunderlich, Purdue University: B.S. in Biochemistry (2017)
Academic Mentor: Adam Siepel

2017 Entering Class: (*Top row,* from *left* to *right*) Yuzhao (Richard) Hu, Sara Boyle, Lyndsey Aguirre, Benjamin Harris. (*Bottom row,* from *left* to *right*) Cole Wunderlich, Diogo Maia e Silva, Dennis Maharjan, Bruno Gegenhuber.

Academic Mentoring

The Watson School takes great pride in the mentoring that it offers its students. One example is our two-tiered mentoring approach, whereby each student chooses both an academic and a research mentor. The academic mentor is a critical advisor during the intensive coursework of the first term, during their rotations, and when identifying a suitable research mentor. Furthermore, the academic mentor continues to follow them throughout their doctoral experience, often serving as important advocates for the students. Entering students select, by mutual agreement, a member of the research or nonresearch faculty to serve as their academic mentor. This program continues to receive much support from the faculty who volunteer to be academic mentors, and it has rightfully become a vital ingredient in our success. The following are the Academic Mentors for the Entering Class of 2017:

STUDENT	MENTOR
Lyndsey Aguirre	Ullas Pedmale
Sara Boyle	Jessica Tollkuhn
Bruno Gegenhuber	John Inglis
Benjamin Harris	W. Richard McCombie
Richard Hu	Jay Lee
Dennis Maharjan	Florin Albeanu
Diogo Maia e Silva	Bruce Stillman
Cole Wunderlich	Adam Siepel

Recruiting Efforts

This year, we once again focused primarily on targeted visits to graduate fairs and minority conferences to identify prospective students for the School. In addition to these visits, the WSBS generated new mailing lists for the distribution of information to top undergraduate science departments around the world. A multiprogram booklet, incorporating the graduate, undergraduate, and postdoctoral programs, was updated for this recruitment season. Additionally, emails were sent to personalized contacts and an electronic mailing list of more than 50,000 individuals who receive information from the Cold Spring Harbor Laboratory Press or have attended Meetings or Courses at the lab. We are grateful to these departments for sharing this contact list. We received 216 applications for the Entering Class of 2018, a slight increase from 2017, and it appears that many outstanding candidates have once again applied to the program.

WATSON SCHOOL OF BIOLOGICAL SCIENCES 2017 RECRUITMENT SCHEDULE

Event	Location	Date
American Association of Cancer Research, Annual Meeting	Washington, D.C.	April 1–5
Cold Spring Harbor Laboratory, Open House	Cold Spring Harbor, New York	June 10
California McNair Scholars Symposium, Graduate School Fair	Berkeley, California	July 27–30
Hunter College MARC and MBRS/RISE, Information Session	New York, New York	September 13
MKN McNair Heartland Research Conference, Graduate School Fair	Kansas City, Missouri	September 22–24
Big 10+ Graduate School Expo, Graduate School Fair	West Lafayette, Indiana	September 24–25
Virtual Graduate School Fair, Hosted by McNairScholars.com and the University of Central Florida Academic Advancement Programs	Online Forum	October 3
Columbia University, Neuroscience Society, Information Session	New York, New York	October 10
Rutgers University: Center for Undergraduate Research, Molecular Biology/Biochemistry Society, and Honors College, Information Session	New Brunswick, New Jersey	October 12

(*continued*)

WATSON SCHOOL OF BIOLOGICAL SCIENCES 2017 RECRUITMENT SCHEDULE (*continued*)

Event	Location	Date
Columbia University, Biology/Biomedical Engineering, Information Session	New York, New York	October 13
American Society for Human Genetics, Annual Meeting	Orlando, Florida	October 17–21
Society for Advancement of Chicanos and Native Americans in Science (SACNAS), National Conference	Salt Lake City, Utah	October 19–21
Notre Dame University, Fall Undergraduate Research Fair, College of Science Research Fair	Notre Dame, Indiana	October 26
University of California, Santa Cruz, Graduate and Professional School Fair	Santa Cruz, California	October 26
Queensborough Community College, Internship and Cooperative Education Forum	Bayside, New York	November 1
Annual Biomedical Research Conference for Minority Students (ABRCMS), National Conference	Phoenix, Arizona	November 1–4
California Forum for Diversity in Graduate Education, Graduate School Fair	Camarillo, California	November 4
New York University, Society for Undergraduate Neural Science, Information Session	New York, New York	November 7
Virtual Summer Research Program Fair Hosted by McNairScholars.com and the University of Central Florida Academic Advancement Programs	Online Forum	November 9
Society for Neuroscience Annual Meeting, Graduate School Fair	Washington, D.C.	November 11–15
University of Maryland, Baltimore County, Meyerhoff Scholars Program, Graduate School Fair	Baltimore, Maryland	November 15
American Society for Cell Biology, Annual Meeting	Philadelphia, Pennsylvania	December 2–6

Students from Other Institutions

WSBS students account for approximately half of the total graduate student population here at CSHL; the other half comprises visiting graduate students from other universities who have decided to conduct some or all of their thesis research in CSHL faculty members' laboratories. A large fraction of students are from Stony Brook University (SBU), via a program established between CSHL and SBU more than 40 years ago. Over the years, we have established relationships with other institutions around the world, enabling their students to conduct research here at CSHL. Currently, we have visiting students from institutions in China, France, Italy, Russia, Slovakia, the Netherlands, and the United States. The Watson School provides a contact person for the students and maintains relationships with the administrators from their home institutions. These students are fully integrated into the CSHL community and receive all the necessary assistance as they navigate the complexities of performing doctoral research away from their home institutions. The following students, listed in the box below, joined us from SBU this year:

STUDENT	CSHL RESEARCH MENTOR	SBU PROGRAM
Jason Carter	Mickey Atwal	Applied Mathematics and Statistics
Kelly Hills-Muckey	Christopher Hammell	Genetics
Shruti Iyer	W. Richard McCombie	Genetics
Devon Lukow	Jason Sheltzer	Genetics
Manojit Mosur Swamynathan	Lloyd Trotman	Molecular and Cellular Biology
Zhe Qian	Nicholas Tonks	Molecular and Cellular Biology
Suelynn Ren	Adam Kepecs	Neuroscience
Deborah Rupert	Stephen Shea	Neuroscience
Danilo Segovia	Adrian Krainer	Molecular and Cellular Biology
Jennifer Thalappillil	David Tuveson	Molecular and Cellular Pharmacology

Graduate Student Symposium

Each year, the students participate in two Graduate Student Symposia held at the Laboratory's Genome Research Center in Woodbury: one in May, the other in October. Each Symposium consists of senior students giving short talks, and coffee breaks and lunch provide opportunities for more informal interactions. The prize for the best talk of the May session was awarded to Daniel Kepple (WSBS, Koulakov lab), and that for the October session was awarded to Nicholas Gallo (SBU, Van Aelst lab), Michael Gutbrod (WSBS, Martienssen lab), and Katie Meze (WSBS, Joshua-Tor lab). We are grateful to the two student chairs—Sashank Pisupati (WSBS) and Padmina Shrestha (SBU)—for their hard work and to WSBS's Kim Geer for helping make the Symposium a great success.

Graduate Student and Postdoctoral Fellow Departures

With each year come not only new arrivals, but also departures. The following graduate students and postdoctoral fellows left the Laboratory during 2017:

POSTDOCTORAL FELLOWS

Michael Campbell	Michael Feigin	Bryan Lanning	Hillary Schiff
Michelle Carmell	Wesley Frey	Michael Ludwig	Damianos Skopelitis
Iok In Chio	Rowan Herridge	Mehrab Modi	Carlos Stahlhut Espinosa
Monika Chugh	Aman Husbands	Shinichi Nakamuta	Ante Tocilj
Sarah Diermeier	Byoung Il Je	Hyun Jae Pi	Cao Xu
Max Doerfel	Steffen Knauer	Arnaud Pommier	Qiangqiang Zhang
Gaofeng Fan	Victoria Kuettner	Nalani Sachan	

GRADUATE STUDENTS

Kush Coshic	Yu-Jui Ho	Onyekachi Odoemene	Nitin Shirole
M. Joaquina Delás Vives	Justus Kebschull	Debjani Pal	Eric Szelenyi
Brad Gulko	Fred Marbach	Annabel Romero Hernandez	Yiyang Wu
Abram Handly-Santana	Maria Nattestad	Lei Sheng	Siwei Zhang

Executive Committee

The School's Executive Committee, in its monthly meetings, provides year-round direction for the School and its students through its invaluable policy recommendations. I wish to thank faculty members Mikala Egeblad, Alexander Krasnitz, Bo Li, David Spector, David Stewart, and Anthony Zador for their service in 2017. I would also like to thank student representatives Dongyan Song (SBU) and Jackie Giovanniello (WSBS), who contributed to discussions and provided useful suggestions and feedback from their colleagues.

The Watson School Continues to Benefit from Generous Benefactors

We are extremely grateful for the generous donors whose one-time gifts or continued support made our 2017 programs possible: the Annenberg Foundation, Lita Annenberg Hazen Foundation, Bristol-Myers Squibb Corporation, Mr. and Mrs. Landon Clay, Lester Crown, the Dana Foundation, Henriette and Norris Darrell, the Samuel Freeman Charitable Trust, the William Stamps Farish Fund, the Genentech Foundation, Mr. and Mrs. Alan Goldberg, Florence Gould Foundation, Gonzalo Río Arronte Foundation, William Randolph Hearst Foundation, Dr. and Mrs. Mark Hoffman, Annette Kade Charitable Trust, Mr. David H. Koch, Mr. and Mrs. Robert D. Lindsay and Family, Mr. and Mrs. David Luke III, Mr. and Mrs. William R. Miller, OSI Pharmaceuticals Foundation, Estate of Edward L. Palmer,

Mr. and Mrs. John C. Phelan, the Quick Family, Estate of Elisabeth Sloan Livingston, the Starr Foundation, the Roy J. Zuckerberg Family Foundation, the Ainslie Foundation, and anonymous donors.

We are also grateful for our endowed lectureships: the John P. and Rita M. Cleary Visiting Lectureship, the George W. Cutting Lectureship, the William Stamps Farish Lectureship, the Martha F. Gerry Visiting Lectureship, the Edward H. Gerry Visiting Lectureship, the Edward H. and Martha F. Gerry Lectureship, the Susan T. and Charles E. Harris Visiting Lectureship, the Klingenstein Lectureship, the Mary D. Lindsay Lectureship, the Pfizer Lectureship, the George B. Rathmann Lectureship, the Seraph Foundation Visiting Lectureship, the Sigi Ziering Lectureship, the Daniel E. Koshland Visiting Lectureship, the Michel David-Weill Visiting Lectureship, and the Fairchild Martindale Visiting Lectureship.

We are also very fortunate to hold a prestigious Ruth L. Krischstein National Research Service Award Predoctoral Training Grant from the National Institutes of Health, National Institute of General Medical Sciences, which was competitively renewed for an additional 5 years in 2017.

Student and Alumni Achievements

To date, 100 students have received their Ph.D. degree from the WSBS. Twenty-six graduates have now secured tenure-track faculty positions. Nine of them have already been promoted to Associate Professor, and this year, Zachary Lippman became the first to be appointed as a full professor. Our graduates have also moved into influential positions in administration, publishing, consulting, and industry. In 2017, Patrick Finigan became a Senior Specialist in Regulatory Affairs CMC Biologics at Merck, Wee Siong (Sho) Goh started an independent fellowship position at the Genome Institute of Singapore, Matt Koh joined Bloomberg as a Natural Language Processing Research Scientist, Colin Malone is now an Assistant Professor and Director of Genomic Analysis and Technical Operations at Columbia University, Katherine McJunkin joined the National Institutes of Health as a Tenure-Track Investigator, Cinthya Zepeda Mendoza is a Laboratory Genetics and Genomics Fellow at the Mayo Clinic, Maria Nattestad joined DNA Nexus as a Scientific Visualization Lead, Michael Pautler became a Research Scientist at the Vineland Research and Innovation Centre in Canada, Amy Rappaport is a scientist at Gritstone Oncology, and Kaja Wasik co-founded a biotech company called Gencove based in New York.

In 2017, our current students and alumni were successful in receiving the following prestigious awards and fellowships:

- WSBS students **David Johnson**, a first-year student, and **Benjamin Berube**, a second-year student in Rob Martienssen's laboratory, were awarded Graduate Research Fellowships from the National Science Foundation. First-year student **Olya Spassibojko** was given honorable mention.
- WSBS graduate **Monica Dus** received an NIH New Innovator award. She also received a 2017 Sloan Research Fellowship in Neuroscience.
- WSBS graduate **Michelle Cilia** received a Presidential Early Career Awards for Scientists and Engineers (PECASE).
- Recent WSBS graduate **Annabel Romero Hernandez**—our resident student artist—had one of her drawings selected by *The Scientist* magazine as their "Image of the Day."

Prizes for the best posters by a postdoctoral fellow and by a graduate student were awarded at the Laboratory's annual In-House Symposium held in November 2017. The poster session provides

2017 WSBS STUDENT (CURRENT OR PREVIOUS) PUBLICATIONS

Albrengues J, Wysocki RW, **Maiorino L**, Egeblad M. 2017. Re-cyclin' cell-cycle components to make NETs. *Dev Cell* **43:** 379–380.

Alexander J, Kendall J, McIndoo J, Rodgers L, **Aboukhalil R**, Levy D, Stepansky A, Sun G, Chobadjiev L, Riggs M, et al. 2017. Utility of single-cell genomics in diagnostic evaluation of prostate cancer. *Cancer Res* **8:** 348–358.

Batut PJ, Gingeras TR. 2017. Conserved noncoding transcription and core promoter regulatory code in early *Drosophila* development. *Elife* **6:** e29005.

Chang AY, **Castel SE**, Ernst E, Kim HS, Martienssen RA. 2017. The conserved RNA-binding cyclophilin, Rct1, regulates small RNA biogenesis and splicing independent of heterochromatin assembly. *Cell Rep* **19:** 2477–2489.

Chen M, Nowak DG, Narula N, Robinson B, Watrud K, Ambrico A, Herzka TM, Zeeman ME, Minderer M, Zheng W, **Ebbesen SH**, et al. 2017. The nuclear transport receptor Importin-11 is a tumor suppressor that maintains PTEN protein. *J Cell Biol* **216:** 641–656.

Chin CS, Peluso P, Sedlazeck FJ, **Nattestad M**, Concepcion GT, Clum A, Dunn C, O'Malley R, Figueroa-Balderas R, Morales-Cruz A, et al. 2017. Phased diploid genome assembly with single-molecule real-time sequencing. *Nat Methods* **13:** 1050–1054.

Delás MJ, Sabin LR, Dolzhenko E, Knott SR, Munera Maravilla E, Jackson BT, Wild SA, Kovacevic T, Stork EM, Zhou M, et al. 2017. lncRNA requirements for mouse acute myeloid leukemia and normal differentiation. *Elife* **6:** e25607.

Delás MJ, Hannon GJ. 2017. lncRNAs in development and disease: From functions to mechanisms. *Open Biol* **7:** 170121.

Faehnle CR*, **Walleshauser J***, Joshua-Tor L. 2017. Multi-domain utilization by TUT4 and TUT7 in control of let-7 biogenesis. *Nat Struct Mol Biol* **4:** 658–665.

Feigin ME, **Garvin T**, Bailey P, Waddell N, Chang DK, Kelley DR, Shuai S, Gallinger S, McPherson JD, Grimmond SM, et al. 2017. Recurrent noncoding regulatory mutations in pancreatic ductal adenocarcinoma. *Nat Genet* **49:** 825–833.

Krug L, Chatterjee N, Borges-Monroy R, Hearn S, Liao WW, Morrill K, Prazak L, Rozhkov N, Theodorou D, Hammell M, Dubnau J. 2017. Retrotransposon activation contributes to neurodegeneration in a *Drosophila* TDP-43 model of ALS. *PLoS Genet* **13:** e1006635.

Kumar V, Rosenbaum J, Wang Z, **Forcier T**, Ronemus M, Wigler M, Levy D. 2017. Partial bisulfite conversion for unique template sequencing. *Nucleic Acids Res* **46:** e10.

Licata AM, Kaufman MT, Raposo D, Ryan MB, **Sheppard JP**, Churchland AK. 2017. Posterior parietal cortex guides visual decisions in rats. *J Neurosci* **37:** 4954–4966.

Manchado E, Huang CH, **Tasdemir N**, Tschaharganeh DF, Wilkinson JE, Lowe SW. 2017. A pipeline for drug target identification and validation. *Cold Spring Harbor Symp Quant Biol* **81:** 257–267.

Öhlund D*, **Handly-Santana A***, Biffi G*, Elyada E*, Almeida AS, Ponz-Sarvise M, Corbo V, Oni TE, Hearn SA, Lee EJ, et al. 2017. Distinct populations of inflammatory fibroblasts and myofibroblasts in pancreatic cancer. *J Exp Med* **214:** 579–596.

Pal D, Pertot A, Shirole NH, Yao Z, Anaparthy N, **Garvin T**, Cox H, Chang K, **Rollins F**, Kendall J, et al. 2017. TGF-β reduces DNA ds-break repair mechanisms to heighten genetic diversity and adaptability of $CD44^{+}/CD24^{-}$ cancer cells. *Elife* **6:** e21615.

Peikon ID*, **Kebschull JM***, Vagin VV, Ravens DI, Sun YC, Brouzes E, Corrêa IR Jr, **Bressan D**, Zador AM. 2017. Using high-throughput barcode sequencing to efficiently map connectomes. *Nucleic Acids Res* **45:** e115.

Roe JS, Hwang CI, Somerville TDD, Milazzo JP, Lee EJ, Da Silva B, **Maiorino L**, Tiriac H, Young CM, Miyabayashi K, et al. 2017. Enhancer reprogramming promotes pancreatic cancer metastasis. *Cell* **170:** 875–888.

Romero Hernandez A, Furukawa H. 2017. Novel mode of antagonist binding in NMDA receptors revealed by the crystal structure of the GluN1-GluN2A ligand-binding domain complexed to NVP-AAM077. *Mol Pharmacol* **92:** 22–29.

Schorn AJ, **Gutbrod MJ**, LeBlanc C, Martienssen, R. 2017. LTR-retrotransposon control by tRNA-derived small RNAs. *Cell* **170:** 61–71.

Soyk S, Lemmon ZH, Oved M, Fisher J, **Liberatore KL**, Park SJ, Goren A, Jiang K, Ramos A, van der Knaap E, et al. 2017. Bypassing negative epistasis on yield in tomato imposed by a domestication gene. *Cell* **69:** 1142–1155.

Underwood CJ, Henderson IR, Martienssen RA. 2017. Genetic and epigenetic variation of transposable elements in *Arabidopsis*. *Curr Opin Plant Biol* **36:** 135–141.

Ziolkowski PA, **Underwood CJ**, Lambing C, Martinez-Garcia M, Lawrence EJ, Ziolkowska L, Griffin C, Choi K, Franklin FC, Martienssen RA, Henderson IR. 2017. Natural variation and dosage of the HEI10 meiotic E3 ligase control *Arabidopsis* crossover recombination. *Genes Dev* **31:** 306–317.

*Authors contributed equally to the work. **Watson School student is designated in boldface.**

WSBS GRADUATES IN FACULTY POSITIONS (IN ORDER OF COMPLETION)

Name	Faculty Position
Amy Caudy	Associate Professor, University of Toronto, Canada
Ira Hall	Associate Professor, Washington University, St. Louis, Missouri
Niraj Tolia	Associate Professor, Washington University, St. Louis, Missouri
Patrick Paddison	Associate Member, Fred Hutchinson Cancer Research Center, Seattle, Washington
Elizabeth Bartom (nee Thomas)	Assistant Professor, Northwestern University, Chicago, Illinois
Michelle Cilia	Research Molecular Biologist, U.S. Department of Agriculture, and Adjunct Assistant Professor, Cornell University, Ithaca, New York
Zachary Lippman	Professor, Cold Spring Harbor Laboratory
Ji-Joon Song	Associate Professor, Korea Advanced Institute of Science and Technology (KAIST), Eoeun-dong, Yuseong-gu, South Korea
Elena Ezhkova	Associate Professor, Mount Sinai School of Medicine, New York, New York
Masafumi Muratani	Associate Professor, University of Tsukuba, Japan
Marco Mangone	Associate Professor, Arizona State University, Tempe
Elizabeth Murchison	Reader, Cambridge University, United Kingdom
Hiroki Asari	Group leader, EMBL Monterotondo, Rome, Italy
François Bolduc	Associate Professor, University of Alberta, Edmonton, Canada
Wei Wei	Assistant Professor, University of Chicago. Illinois
Christopher Harvey	Assistant Professor, Harvard University, Cambridge, Massachusetts
Tomas Hromadka	Project Leader, Slovak Academy of Sciences, Bratislava, Slovakia
Monica Dus	Assistant Professor, University of Michigan, Ann Arbor
Daniel Chitwood	Assistant Professor, Donald Danforth Plant Science Center, St. Louis, Missouri
Jeremy Wilusz	Assistant Professor, University of Pennsylvania, Philadelphia
Colin Malone	Assistant Professor and Director of Genomic Analysis and Technical Operations, Columbia University, New York
Oliver Fregoso	Assistant Professor, University of California, Los Angeles
Hiroshi Makino	Assistant Professor, Nanyang Technological University, Singapore
Katherine McJunkin	Tenure Track Investigator, National Institutes of Health, Bethesda, Maryland
Yaniv Erlich	Assistant Professor, New York Genome Center, Columbia University, New York
Michael Pautler	Research Scientist, Vineland Research and Innovation Centre, Ontario, Canada

WSBS GRADUATES IN NONACADEMIC POSITIONS (IN ORDER OF COMPLETION)

Name	Current Position
Emiliano Rial-Verde	Director, Corporate Development, Bunge Limited, White Plains, New York
Rebecca Ewald	International Business Leader, Ventana Medical Systems/Roche, Tucson, Arizona
Catherine Seiler (née Cormier)	Manager, The Biobank Core Facility, St. Joseph's Hospital and Medical Center, Phoenix, Arizona
Darren Burgess	Senior Editor, *Nature Reviews Genetics*, United Kingdom
Rebecca Bish-Cornelissen	Senior Scientific Editor, D.E. Shaw Research, New York
Allison Blum	Science Operations Manager, HHMI/Columbia University, New York
Keisha John	Director of Diversity Programs, University of Virginia, Charlottesville
Oliver Tam	Sequencing Analyst, ARUP Laboratories, Salt Lake City, Utah
Amy Rappaport	Senior Scientist, Gristone Oncology, Emeryville, California
Frederick Rollins	Engagement Manager, LEK Consulting, Boston, Massachusetts
Patrick Finigan	Senior Specialist, Regulatory Affairs CMC Biologics, Merck
Elizabeth Nakasone	Resident Physician, UCLA Health
Maria Pineda	Co-Founder, CEO, Envisagenics, New York, New York
Felix Schlesinger	Bioinformatics Scientist, Illumina, Inc., San Diego, California
Paloma Guzzardo	Team Leader, Horizon Discovery, Vienna, Austria
Saya Ebbesen	Senior Medical Writer, BluPrint Oncology, London, United Kingdom
Joshua Sanders	Founder and C.E.O., Sanworks, L.L.C., Stony Brook, New York

(*continued*)

WSBS GRADUATES IN NONACADEMIC POSITIONS (IN ORDER OF COMPLETION) (*continued*)

Name	Current Position
Kaja Wasik	Co-Founder, Gencove, New York
Mitchell Bekritsky	Bioinformatic Scientist, Illumina, Inc., Cambridge, United Kingdom
Sang-Geol Koh	Scientist and Entrepreneur, {Mind}, South Korea
Susann Weissmueller	Strategic Partnering Associate, Roche, Switzerland
Ian Peikon	Lead Scientist, Kallyope, New York, New York
Cinthya Zepeda Mendoza	Laboratory Genetics and Genomics Fellow, Mayo Clinic, Boston, Massachusetts
Jack Walleshauser	Research Scientist at Plexxikon Inc., Berkeley, California
Lisa Krug	Scientist, Kallyope, New York, New York
Robert Aboukhalil	Senior Bioinformatics Software Engineer, GenapSys, Redwood City, California
Tyler Garvin	Senior Software Engineer, Pict, California
Anja Hohmann	Senior Scientist, KSQ Therapeutics, Boston, Massachusetts
Charles Underwood	Scientist, KeyGene, Wageningen, Netherlands
Matt Koh	Natural Language Processing Research Scientist, Bloomberg LP, New York, New York
Annabel Romero Hernandez	Associate Scientist, Regeneron Pharmaceuticals, Tarrytown, New York
Maria Nattestad	Scientific Visualization Lead, DNA Nexus and Founder, OMGenomics

a forum for the postdoctoral fellows and students to show off their research and gives the entire scientific community a chance to come together and share ideas beyond the walls of their individual laboratories. This year, Emilis Bruzas, a Watson School student from Mikala Egeblad's laboratory, and Anqi Zhang, a Watson School student from Tony Zador's laboratory, shared the graduate student prize. The postdoctoral prize was shared by Grinu Mathew from Lloyd Trotman's laboratory and Anirban Paul from Josh Huang's laboratory.

Alexander Gann
WSBS Professor and Dean

SPRING CURRICULUM

TOPICS IN BIOLOGY

Each year, invited instructors offer 7-day courses at the Banbury Conference Center exploring specialized topics outside the expertise of the Cold Spring Harbor Laboratory faculty. These courses include morning and evening lectures, as well as afternoon sessions during which students read assigned papers or work on problem sets and presentations. In spring 2017, there were two courses: Immunology and Physical Biology of the Cell.

Immunology

February 5–11	**Attended by the entering classes of 2013 and 2016**
INSTRUCTOR	**Hidde Ploegh,** Boston Children's Hospital
TEACHING FELLOWS	**Carlos Donado,** Harvard University **Martin Fan,** Harvard University **Ayano Kohlgruber,** Harvard University

Immunology focuses on understanding the mechanisms by which multicellular organisms defend themselves against external threats of microbial aggression and internal threats associated with genetic instability and cellular transformation. The course focused on the innate immune system and the adaptive immune system. Innate immunity defends against microbes by recognizing evolutionarily conserved molecular patterns; the adaptive immune system has enormous flexibility in molecular recognition, but it can also target self to cause autoimmune diseases.

Physical Biology of the Cell

March 12–18	**Attended by the entering classes of 2014 and 2015**
INSTRUCTOR	**Rob Phillips,** Caltech
VISITING LECTURERS	**Jané Kondev,** Brandeis University **Hernan Garcia,** University of California, Berkeley **Madhav Mani,** Northwestern University **Aleksandra Walczak,** Ecole Normale Supérieure
TEACHING FELLOWS	**Muir Morrison,** Caltech

The aim of this course was to provide a hands-on experience in the use of quantitative models as a way to view biological problems. They began with "order of magnitude biology," showing how simple estimates can be exploited in biology. They showed how to construct simple models of a variety of different biological problems, primarily using the tools of statistical mechanics. One of the key themes of the course was to show how physical biology unites and organizes topics in a

fundamentally different way, often revealing how topics that are nearby in physical biology seem unrelated when viewed from the vantage point of molecular or cell biology. The instructors guided the students from start to finish on several modeling case studies.

SPECIAL COURSES

Optical Methods

January 5–7 **Attended by the entering class of 2016**

INSTRUCTOR **Florin Albeanu,** Cold Spring Harbor Laboratory

TEACHING FELLOWS **Walter Bast,** Cold Spring Harbor Laboratory
Quentin Chevy, Cold Spring Harbor Laboratory

Optical imaging techniques are widely used in all areas of modern biological research. Our aim for this course was to give students an introduction into widely used basic and advanced optical methods. Given the experimental nature of the topic, a central aim of our course was to offer students a practical hands-on experience. This included both the use of commercially available systems and, more importantly, a primer on custom building and adapting optical setups to address specific biological needs.

Teaching Experience at the Dolan DNA Learning Center

Entering Class of 2016

DIRECTOR **David A. Micklos**

INSTRUCTORS **Amanda McBrien** (Lead)
Elna Gottlieb
Katie McAuley
Erin McKechnie
Bruce Nash
Sharon Pepenella

As science plays an increasing role in society, there is also an increasing need for biologists to educate nonscientists of all ages about biology. The Watson School of Biological Sciences doctoral program offers its students the opportunity to teach in the Laboratory's Dolan DNA Learning Center, where students teach laboratory courses to high school and middle school students. In so doing, they learn how to communicate with nonbiologists and to inspire and educate creative young minds. The teaching module entailed pairs of students teaching one morning or afternoon a week for 12 weeks. In the initial weeks, the Dolan DNA Learning Center instructors taught

the Watson School students the didactic process; it was not until the fifth week that the graduate students taught on their own. At the end of the 12 weeks, the students were very excited about their teaching experience.

LABORATORY ROTATIONS

Entering Class of 2016

ROTATION MENTORS	**Mickey Atwal**	**Jesse Gillis**	**Zachary Lippman**	**Nick Tonks**
	Anne Churchland	**Chris Hammell**	**Rob Martienssen**	**David Tuveson**
	Camila dos Santos	**Leemor Joshua-Tor**	**Alea Mills**	**Chris Vakoc**
	Mikala Egeblad	**Adrian Krainer**	**Ullas Pedmale**	**Linda Van Aelst**
	Doug Fearon	**Jay Lee**	**Stephen Shea**	**Doreen Ware**
	Hiro Furukawa	**Bo Li**	**Adam Siepel**	**Tony Zador**

The most important element of a doctoral education is learning to perform independent research. After the fall term courses, students participate in laboratory rotations; these provide students and faculty the opportunity to get to know each other and to explore possibilities for doctoral thesis research. At the end of each rotation, students make short presentations of their studies to the other students and their rotation advisors. These talks give students an opportunity to share their laboratory experiences and to practice giving scientific presentations. This year, 24 WSBS faculty members served as rotation mentors, some mentoring more than one student.

FALL CURRICULUM

ENTERING CLASS OF 2017

The students started the semester by attending boot camps in Molecular, Cellular, and Quantitative Biology, as an introduction to the techniques and terminology that they would encounter in subsequent courses. The Molecular and Cellular Biology boot camp featured seven lectures from faculty members Florin Albeanu, Hiro Furukawa, Dick McCombie, Pavel Osten, Lloyd Trotman, Linda Van Aelst, and Lingbo Zhang, and the Quantitative Biology boot camp lectures were given by Mickey Atwal and Justin Kinney.

CORE COURSES

The Leslie C. Quick, Jr. Core Course on Scientific Reasoning and Logic

INSTRUCTORS	**Linda Van Aelst (Lead)**	**Leemor Joshua-Tor**
	Alexander Gann	**Bo Li**
	Christopher Hammell	
GUEST LECTURERS	**Hiro Furukawa**	**Adrian Krainer**
	Z. Josh Huang	**Robert Martienssen**
	Justin Kinney	**Christopher Vakoc**
TEACHING ASSISTANTS	**Sonali Bhattacharjee**	**Jon Ipsaro**
	Cristina Chen	**Luis Mejia**

In this core course, which forms the heart of the curriculum, students (1) acquired a broad base of knowledge about the biological sciences, (2) learned the scientific method, and (3) learned how to think critically. The initial four to five modules were on a different general theme; in each module, students read an assigned set of research articles, and at the end of the module, they were provided written answers to a problem set that guides them through several of the articles.

Twice weekly, students attended lectures related to the module's topic that included concepts and fundamental information as well as experimental methods. The students met among themselves to discuss the assigned papers not covered by the problem set. Each week, students spent an evening discussing the assigned articles with faculty. In the final module of the course, students participated in a mock study section in which real National Institutes of Health R01 grants were reviewed and critiqued. This allowed students to evaluate the questions before the answers were known, evaluate routes toward discovery before knowing where they will end, and make critical judgments about how to proceed in the face of an uncertain outcome.

In 2017, the following were the module topics for this course:

Topic	Instructor(s)
Gene Expression	Alex Gann
Gene Regulatory Logic and the Construction of Multicellular Organisms: Insights from Flies, Plants, and Worms	Christopher Hammell
The Brain: Wiring, Plasticity, and Maladaptation	Bo Li
Macromolecular Structure and Function	Leemor Joshua-Tor
Study Section	Linda Van Aelst

The Darrell Core Course on Scientific Exposition and Ethics

The 2017 Scientific Exposition and Ethics (SEE) core course was revamped this year with three distinct sections covering writing, oral communication, and ethics. It was taught by lead instructor Mikala Egeblad, along with David Jackson, Sydney Gary, and Charla Lambert. As usual, the course hosted expert guest lecturers who covered topics including scientific presentations, ethical and legal responsibilities of scientists, DNA profiling and postconviction appeals, and reproductive genetics.

INSTRUCTORS	**Mikala Egeblad (Lead)**	**David Jackson**
	Sydney Gary	**Charla Lambert**
GUEST LECTURERS	**Lisa Bianco**	**Ullas Pedmale**
	Diane Esposito	**Richard Sever**
	Jonathan Ipsaro	**Rachel Strittmatter**
	Alyson Kass-Eisler	

VISITING LECTURERS

Keith Baggerly, M.D. Anderson Cancer Center
Robert Charrow, Greenberg Traurig, LLC
Graham Chedd, Alan Alda Center for Communicating Science
Susan Friedman, The Innocence Project
Avner Hershlag, North Shore University Hospital
David Miller, Graphic Designer

This core course offered instruction in the fundamental elements of scientific exposition—writing skills and public speaking—and ethics. The ability to communicate effectively and to appreciate the intricacies of ethical issues are essential skills for biologists; both subjects were taught in a series of example-based lectures and discussion groups. Writing skills included the fundamentals of modern scientific English and the organization and preparation of papers, research abstracts, and grant applications. Oral presentation skills were taught by instructors with different modes of presentation. Together with instructors, students critiqued formal seminar presentations at the Laboratory. Instruction and discussions about ethics included the ethical implications of biological discovery for society as well as the nature and boundaries of ethical behavior of scientists and their rights and responsibilities. A primary objective of the course was that students consider exposition and ethics integral aspects of scientific research.

Research Topics

ORGANIZERS **Kimberley Geer**
Alyson Kass-Eisler

This core course provided students with an in-depth introduction to the fields of research that the Laboratory scientists investigate. Students and faculty attended a weekly Research Topics seminar, at which faculty members and CSHL fellows presented their current research topics and methods of investigation each Wednesday evening over dinner. The students learned how to approach important problems in biology. These seminars, together with the annual fall Laboratory In-House Symposium, provided students with a basis for selecting laboratories in which to do rotations.

SPECIALIZED DISCIPLINES COURSES

The students in the Entering Class of 2017 took a total of four Specialized Disciplines courses this fall: *Quantitative Biology, Genetics and Genomics, Cancer,* and *Systems Neuroscience.*

Quantitative Biology

Throughout the semester

INSTRUCTOR **Adam Siepel (Lead)**

CSHL GUEST LECTURERS **Mickey Atwal** **Alexander Krasnitz**
Justin Kinney **Dan Levy**

Quantitative reasoning is a powerful tool to uncover and characterize biological principles, ranging from the molecular scale all the way to the ecological. With the advent of high-throughput technologies in genomics and neuroscience, it has become increasingly necessary for biological researchers to be able to analyze and interpret high-dimensional data and frame biological hypotheses mathematically. To this end, this course aimed to equip the students with a basic training in computer programing, modern statistical methods and physical biology. By the end of the course, not only were the students able to answer many of the statistical questions that arise in data analyses, but they also became familiar with the more complex techniques used by fellow computational biologists. Topics covered included probabilities, statistical fluctuations, Bayesian inference, significance testing, fluctuations, diffusion, information theory, neural signal processing, dimensional reduction, Monte Carlo methods, population genetics, and DNA sequence analyses. A common theme throughout the course was the use of probabilistic and Bayesian approaches.

Genetics and Genomics

September 5–28

INSTRUCTOR	**Zachary Lippman (Lead)**
GUEST LECTURER	**Ullas Pedmale**
VISITING LECTURERS	**Guillaume Lettre (University of Montreal)**

This course placed modern genetics and genomics into the context of classical genetics. History, technique, and perspective of genetic inference were described around four levels of analysis: forward genetics, natural genetic variants, gene interaction, and genomics. Emphasis was placed on integrating classical with modern questions of genetic analysis: How are genes mapped and "cloned"? How do gene mutations help to define biological processes? How are more complex traits genetically dissected into simpler (underlying) components? What concepts and techniques are used to organize genes into pathways and networks? What defines a gene and what gene variation exists in natural populations? What are the functional consequences of gene variation, and how is it detected? How are genomes organized and coordinately regulated? How can genomic information be cataloged, organized, and mined? These questions and concepts were fleshed out using examples from the literature.

Cancer

October 10–24

INSTRUCTOR	**David Tuveson (Lead)**	
GUEST LECTURERS	**Mikala Egeblad**	**Bruce Stillman**
	Molly Hammell	**Nicholas Tonks**
	Jay Lee	**Lloyd Trotman**
	David Spector	**Christopher Vakoc**
TEACHING ASSISTANT	**Lindsey Baker**	

Cancer represents an increasing cause of morbidity and mortality throughout the world as health advances continue to extend the life spans of our populations. Although our basic understanding of cancer has increased considerably since 1971, when United States President Richard Nixon initiated the "War on Cancer," our ability to translate this knowledge into a health benefit for patients has been restricted to certain malignancies and often only temporarily. Importantly, specific hypotheses developed from our knowledge of cancer biology can be tested in increasingly complex model systems ranging from cell culture to genetically engineered mouse models, and such investigations should prove invaluable in discovering new methodologies for the detection, management, and treatment of cancer in humans.

At the conclusion of this course, students were able to elaborate an understanding of cancer as a pathobiological process that invades our bodies without offering any known benefit to the host,

discuss how we diagnose cancer today, and contemplate how to replace the methods currently used to treat cancer. Students were also able design tractable methods to investigate fundamental aspects of cancer biology, and became familiar with translational approaches to defeating cancer. Topics covered in this course included biochemistry, epigenetics, immunology, resistance, growth control, microenvironment, noncoding RNA, and disease modeling. The implications of the biological findings for cancer prevention, diagnosis, and treatment were covered.

Systems Neuroscience

October 23–November 17

INSTRUCTORS	**Stephen Shea (Lead)** **Florin Albeanu**
GUEST LECTURERS	**Adam Kepecs** **Jessica Tollkuhn**
TEACHING ASSISTANT	**Priyanka Gupta**

This course provided an overview of key aspects of neuroscience, with a focus on learning and plasticity from its cellular basis, through development, to systems and behavior. Both experimental and theoretical viewpoints were explored.

POSTDOCTORAL PROGRAM

PROGRAM DIRECTOR **Nicholas Tonks**

PROGRAM ADMINISTRATOR **Alyson Kass-Eisler**

An important measure of our postdoctoral program's success is the ability of postdoctoral fellows to secure positions after they complete their training. This year, our fellows accepted positions at Weill Cornell Medicine; Chinese Academy of Sciences; ENZO Life Sciences; Yonsei University in Seoul, South Korea; New York Institute of Technology; Roswell Park Cancer Center; 10x Genomics; the Sainsbury Wellcome Centre, United Kingdom; Shanghai Tech University, China; and University of Otago, New Zealand.

The **Postdoctoral Liaison Committee** (PDLC), which is an elected group of postdoctoral fellows who communicate information and ideas between the administration and the postdoctoral community, continues to enhance CSHL's postdoctoral experience. The PDLC is essentially the voice of the community and holds regular meetings with Dr. Bruce Stillman, CSHL President. In 2017, new member Michael Campbell joined returning PDLC members Cristina Aguirre-Chen, Sonali Bhattacharjee, Dhananjay Huilgol, Grinu Matthew, and Michael Regan. The PDLC organized a successful daylong retreat this year aimed at fostering networking and collaboration. The retreat took place at the Banbury Conference Center and included panels on academic and nonresearch careers, research presentations by current postdocs, a "mock chalk talk," and an Improv Workshop.

The PDLC, along with the CSHL Library, continued the very successful Perspectives on Science Careers series that was started last year. This year, the guest speakers included Colleen Cuffaro from Canaan Partners on Careers in venture capital and equity research; Jonathan Leff, Robert Jackson, William Slattery, and Karen Heidelberger from the Deerfield Institute on healthcare venture; Brandy Houser from Celdara Medical on transitioning from bench science to the business of science; and Pavel Khrimian from MedImmune (AstraZeneca) on careers in pharma.

The PDLC also oversees and distributes funds provided by Dr. Stillman to two career development groups: the Career Development Program and the Bioscience Enterprise Club. These groups are primarily composed of postdoctoral fellows, but they also include graduate students. Today's postdoctoral fellows face a number of challenges, including a very difficult and competitive job market. CSHL endeavors to prepare postdocs to be competitive for the limited number of jobs available. It is increasingly becoming CSHL's role to introduce the diversity of career opportunities available and to provide the tools postdocs need to prepare for these positions. As a result, a number of events were organized with the assistance of the PDLC and career development groups.

Career Development Program

The Career Development Program (CDP) provides programming geared toward careers in academia. The successful Conversations with Faculty lecture series, where postdocs are provided career insights within an informal and interpersonal format, was reimagined as Coffee Chat. The new format was initiated by new faculty member Ullas Pedmale and was coorganized by postdoc Leah Banks. The 2017 sessions were hosted by faculty members Leemor Joshua-Tor, David Spector, Stephen Shea, David Jackson, and new Associate Dean Monn Monn Myat.

CDP hosted a special lecture on "How to Prepare and Give an Effective Chalk Talk," given by Dr. Leslie Vosshall, Professor, the Rockefeller University. The "Chalk Talk," an essential part of

an academic job interview, is one of the hardest talks for which to prepare. Candidates are given a chance to present their future research goals without the use of visual aids. Without guidance, preparation, and practice, these talks can be difficult to give and can be the difference in getting a faculty position. A mock chalk talk was also presented by a CSHL postdoc at the PDLC retreat.

The CDP has also been bolstering their connections with local colleges and universities in order to provide teaching opportunities for interested CSHL postdocs. Informal partnerships have now been established with Adelphi University, Molloy College, and Long Island University. In addition, selected postdoctoral fellows continue to participate in the Watson School as tutors (either one-on-one or in the classroom setting), thus providing CSHL postdocs with valuable experience in teaching and mentoring.

Bioscience Enterprise Club

The Bioscience Enterprise Club (BEC) disseminates information about nonacademic careers to the CSHL postdoc community. Topic areas include biotechnology, intellectual property, law, regulatory affairs, and venture capitalism. This year, the BEC organized their biennial, day-long, "Beyond the Bench Symposium," which included sessions on the following topics:

Careers in Education and Government featuring Jason Williams, Assistant Director, DNA Learning Center; Jaclyn Tetenbaum-Novatt, Assistant Professor, Long Island University; Devinn Lambert (WSBS Alumnus), Technology Manager, U.S. Department of Energy; and Julie Wu, Primary Patent Examiner, U.S. Patent and Trademark Office.

Careers in Pharma and Biotech featuring Marc Hild, Senior Investigator, Novartis Institutes for Biomedical Research; Qing Xiang, Senior Principal Scientist, Pfizer; Anja Hohmann (WSBS Alumnus), Senior Scientist, KSQ Therapeutics; and Tamsen Dunn, Bioinformatics Scientist, Illumina.

Careers in Start-ups and Data Science featuring Ian Peikon (WSBS Alumnus), Senior Scientist, Kallyope; Kaja Wasik (WSBS Alumnus), Co-Founder, Gencove; Daniel Ferrante, Co-Founder and CSO, SFL Scientific; and Matthew Koh (WSBS Alumnus), Insight Data Science Fellow.

Careers in Science Management, Law, and Finance featuring Carmella Stephens, Patent Attorney and Partner, CDFS Law; Andrew Whiteley, Director of Scientific and Medical Programs, The Lustgarten Foundation; Sam Hall, Principal, Apple Tree Partners; and Priya Shridevi, Research Investigator, CSHL.

The keynote speaker was Jennifer Doudna, a Professor at the University of California, Berkeley, and the cofounder of Caribou Biosciences.

Along with the Women in Science and Engineering (WiSE) group at CSHL, BEC hosted a workshop on "How to Negotiate for What You Need" by Dr. Sandra Masur, Director of the Office for Women's Careers at the Icahn School of Medicine at Mount Sinai.

BEC also organized a trip to Regeneron Pharmaceuticals that enabled postdocs to participate in the Regeneron Science to Medicine Forum.

National Postdoc Appreciation Week

An ice cream social was held on CSHL's campus to celebrate National Postdoc Appreciation week. These events provided a great opportunity for the community to join together for some fun, while networking and learning about CSHL's ongoing programs.

In-House Symposium

Prizes for the best posters by a postdoctoral fellow and by a graduate student were awarded at the Laboratory's annual In-House Symposium held in November 2017. The poster session provides a

forum for the postdoctoral fellows and students to show off their research and gives the entire scientific community a chance to come together and share ideas beyond the walls of their individual laboratories. This year, Emilis Bruzas, a Watson School student from Mikala Egeblad's laboratory, and Anqi Zhang, a Watson School student from Tony Zador's laboratory, shared the graduate student prize. The postdoctoral prize was shared by Grinu Mathew from Lloyd Trotman's laboratory and Anirban Paul from Josh Huang's laboratory.

The Science Alliance

All CSHL postdoctoral fellows and graduate students are enrolled in a special initiative of the New York Academy of Science (NYAS), the Science Alliance. The Science Alliance for graduate students and postdoctoral fellows is a consortium of universities, teaching hospitals, and independent research facilities in the New York City metro area that have formed a partnership with NYAS. The Alliance's aim is to provide career and professional development mentoring for postdoctoral fellows and graduate students in science and engineering, through a series of live events and a dedicated web portal. In addition, the Science Alliance offers graduate students and postdoctoral fellows the opportunity to network with their peers across institutions and with key leaders in academia and industry.

The Science Alliance meetings, courses, and workshops this year discussed the following topics: Clinical Research Management (Online Course); Scientists Teaching Science (Online Course); Risky Business: The Future of Biopharmaceutical Innovation; Learn Basic Computing Skills to Be More Effective in the Lab; Non-Academic Career Tracks for International Scientists: You, Too, Can Peer Review!; What International Scientists Should Know About Immigration and Travel (Webinar); Industry Research Postdoctoral Programs; Securing Seed Funding: Support for Biotech Entrepreneurs and Researchers; Academia Challenges for Women in STEM: Training, Discrimination, and Policy; Science Alliance Leadership Training (SALT); Grantsmanship for Postdocs: Pathway to Independence Awards (K99/R00); SciPhD: The Business of Science Certificate Program; Planning for a Successful STEM Career; and Engaging with the Public: Why Scientists Need to Communicate Their Work.

UNDERGRADUATE RESEARCH PROGRAM

PROGRAM DIRECTORS: **Anne Churchland**
Christopher Hammell

PROGRAM ADMINISTRATOR: **Kimberly Creteur**

Established almost 60 years ago, the CSHL Undergraduate Research Program (URP) provides undergraduates from around the world with hands-on undergraduate research training in biology. The 10-week program begins the first week of June. Several activities are implemented to ensure that URP participants transition smoothly into the Laboratory community and research. For example, during the first week, the students attend various orientations and receive a guided historical tour of campus and all the facilities and resources available to them. The URP attendees work, live, eat, and play among CSHL scientists, and they have a very busy academic and social calendar throughout the remaining 9 weeks of the summer. The students receive training in scientific research, science communication, career preparation, and bioinformatics and computational biology, all while interacting socially with fellow program participants and members of the CSHL community in formal and informal activities. Some of the 2017 activities included a pizza party with Dr. and Mrs. Watson, dinner with Dr. and Mrs. Stillman, volleyball games, designing the URP T-shirt, a Broadway show, and the ever-famous URP versus PI volleyball match.

The students' scientific development is the most important component of the program. At the beginning of the summer, the URP attendees write an abstract and present a talk on their proposed research. The URP participants work alongside scientists and become increasingly independent throughout the summer. Concluding the program in August, the URP students prepare a final report and present their results in a 15-minute talk at the URP Symposium. As in previous years, the program directors and faculty mentors were highly impressed with the accomplishments of the URP students.

The following 20 students, selected from 848 applicants, took part in the 2017 program:

George Bekheet
Advisor: **Churchland Laboratory**
Funding: **National Science Foundation Scholar**
Multisensory integration during an audiovisual looming stimulus.

Isaac Cohen
Advisor: **Huang Laboratory**
Funding: **National Science Foundation Scholar**
Exploring the cellular basis of the circuitry and functional organization of mouse motor cortex.

Sean Connelly
Advisor: **Mills Laboratory**
Funding: **Robert H.P. Olney Fellow/Garfield Fellowship**
LSH, EZH2, and the cancer stem cell population.

Josephine Cooke
Advisor: **Albeanu Laboratory**
Funding: **The City University of New York (CUNY) URP Scholar**
Discrimination of olfactory stimuli in mice using a two-alternative forced-choice paradigm.

Patrick Cunniff
Advisor: **Jackson Laboratory**
Funding: **University of Notre Dame URP Scholar**
Intercellular signaling and transport in *Arabidopsis thaliana*.

Alex Francette
Advisor: **dos Santos Laboratory**
Funding: **William Townsend Porter Foundation Scholar**
Elucidating the role of MLL-3 in altering the parous epigenetic landscape.

Marianna Frey
Advisor: **Kepecs Laboratory**
Funding: **National Science Foundation Scholar**
Adenosinergic modulation of optimal foraging decisions.

James Gornet
Advisor: **Osten Laboratory**
Funding: **National Science Foundation Scholar**
Mapping single neurons from whole-brain images.

Alexander Kirschner
Advisor: **Joshua-Tor Laboratory**
Funding: **Alfred L. Goldberg Fellowship/Joan Redmond Read Fellowship**
Mechanism of epigenetic control by heterochromatin protein 1 (HP1) and origin recognition complex (ORC 2/3).

Likhitha Kolla
Advisor: **Atwal Laboratory**
Funding: **National Science Foundation Scholar**
Mapping the immune landscape for breast cancer subtypes.

Natasa Kostic
Advisor: **Krasnitz Laboratory**
Funding: **National Science Foundation Scholar**
A computational pipeline for absolute copy number quantification in single cancer cells.

Asad Lakhani
Advisor: **Egeblad Laboratory**
Funding: **James D. Watson Undergraduate Scholar**
Existence of a negative feedback loop between LOX and Ras signaling in PDAC.

Yutong (Alyssia) Liu
Advisor: **Trotman Laboratory**
Funding: **Von Stade Fellowship/Burroughs Wellcome Fellowship**
Validating the biology and evolution of genome duplication in prostate cancer.

Fotini Papaleonidopoulos
Advisor: **Krainer Laboratory**
Funding: **James D. Watson Fellow/Dorcas Cummings Scholar**
Antisense-mediated inhibition of nonsense-mediated mRNA decay of *CFTR* gene.

Vir Patel
Advisor: C. **Hammell Laboratory**
Funding: **Libby Fellowship/William Shakespeare Fellowship**
The regulatory protein PQN-59 forms amyloid aggregates dependent on its domain architecture.

Charles Pei
Advisor: **Lippman Laboratory**
Funding: **National Science Foundation Scholar**
Modification of meristem and floral development genes in *Physalis peruviana*.

Dawn Truong
Advisor: **Gingeras Laboratory**
Funding: **National Science Foundation Scholar**
Response of normal mouse cells to mouse tumor-derived extracellular vesicles.

George Wang
Advisor: **Ware Laboratory**
Funding: **National Science Foundation Scholar**
Resources for identifying the genetic basis of important traits in grapes.

Shenandoah Wrobel
Advisor: **Shea Laboratory**
Funding: **National Science Foundation Scholar**
Granule cell modulation of odor representations in awake mice.

Ray Zhang
Advisor: **Stillman Laboratory**
Funding: **Former URP Fund Scholar/30th Anniversary URP Scholar**
Screening domains of CDC6.

SUMMER RESEARCH INTERNSHIP FOR MEDICAL STUDENTS

PROGRAM ADMINISTRATOR **Jessica Gotterer**

Through the CSHL and Northwell Health affiliation, a summer internship program has been created to give students entering into their second year of medical school from the Donald and Barbara Zucker School of Medicine at Hofstra/Northwell the opportunity to spend a summer experiencing basic research in a CSHL lab and attending relevant seminars. To date, students have been offered positions in labs focusing on cancer and the genetics of human disease. Students commit 8–10 weeks (roughly June–August) during the summer following their first year of course work to full-time research in a CSHL lab. The students complete a research project and present their work at the annual "Scholarship Day" at Zucker School of Medicine the following fall.

The following students took part in the 2017 program:

STUDENT	CSHL MENTOR
Emily Cen	Mickey Atwal
Vincent D'Andrea	Lloyd Trotman
Christine Grosso	Mickey Atwal

PARTNERS FOR THE FUTURE

PROGRAM DIRECTOR **David Jackson**

PROGRAM ADMINISTRATOR **Lauren Warmbrand**

The Partners for the Future Program, established by Dr. James Watson in 1990, provides an opportunity for talented Long Island high school students to have hands-on experience in biomedical research at Cold Spring Harbor Laboratory. This highly competitive program is open to Long Island high school students in their junior year. Each high school science chairperson may nominate up to three students. The top candidates are interviewed by CSHL scientists. Students selected for the program are paired with a scientist mentor and spend a minimum of 10 hours per week, September through March of their senior year, conducting original research. At the conclusion, the students present their projects to an enthusiastic audience of the students, scientific mentors and colleagues, CSHL administrators, parents, and teachers. Although the students learn a great deal about modern biology and state-of-the-art research techniques, the main advantage of the program is that they are exposed to day-to-day life in a lab. Interacting with scientists and support staff, the students learn to define and pursue a research goal while solving problems that may occur along the way.

The following 2017–2018 Partners for the Future were chosen from among 72 nominations:

PARTNER	HIGH SCHOOL	LABORATORY	MENTOR
Thomas Carey	Cold Spring Harbor High School	Adam Kepecs	James Sturgill
Katelyn Dorane	Plainedge High School	Doreen Ware	Lifang Zhang
Daniel Lee	Commack High School	Tatiana Engel	Tatiana Engel
Casey Leimboch	Sacred Heart Academy	Robert Martienssen	Sonali Bhattacharjee
Kelsey Nathan	Friends Academy	Stephen Shea	Alexandra Nowlan
Julia Park	Plainview Old-Bethpage JFK High School	Mikala Egeblad	Emilis Bruzas
Nivetha Shanmugarajah	New Hyde Park Memorial H.S.	Ullas Pedmale	Louise Noren Lindback
Paige Silverstein	Friends Academy	David Tuveson	Luana Guerriero
Jerinna Solages	Amityville Memorial High School	Anne Churchland	Lital Chartarifsky
Mahon Walsh	Saint Anthony's High School	Lloyd Trotman	Alexandra Ambrico
Yingdong (Tony) Wen	Long Island Lutheran High School	David Jackson	Edgar Demesa Arevalo
Monet Yuan	Syosset High School	Jason Sheltzer	Jason Sheltzer

MEETINGS & COURSES PROGRAM

ACADEMIC AFFAIRS

The Cold Spring Harbor Laboratory Meetings & Courses program attracts scientists from all over the world to communicate, learn, and critique the latest ideas and approaches in the biological sciences. The Meetings & Courses program at the Laboratory attracted strong attendance in 2017, with 7300 meeting participants and more than 1400 course participants (trainees, teaching, and support faculty). The Cold Spring Harbor Asia program included 19 conferences and attracted more than 3300 participants, bringing the anticipated year-end total for both the United States– and China-based programs to almost 12,000.

The Laboratory held more than 30 academic meetings this year, which brought together scientists from around the world to discuss their latest research. The spring meeting season culminated in the 82nd Cold Spring Harbor Symposium, which focused on Chromosome Segregation and Structure and addressed the enormous progress in our understanding of chromosome segregation, dynamics, and stability. The Symposium attracted almost 300 participants, including notable scientists such as Geneviève Almouzni, Angelika Amon, Marisa Bartolomei, Mónica Bettencourt-Dias, Gerd Blobel, Xin Chen, Don Cleveland, Bernard De Massy, Daniel Durocher, Susan Gasser, Stephen Harrison, Edith Heard, Steven Henikoff, Scott Keeney, Barbara Meyer, Kim Nasmyth, David Page, David Pellman, Jan-Michael Peters, Oliver Rando, Allan Spradling, Stephen West, and Yukiko Yamashita, to name but a few. Dissemination includes the proceedings of the Symposium, published each year by the CSHL Press. Videotaped interviews with leading speakers conducted by editors and journalists attending the Symposium are now available on our Leading Strand YouTube channel. The Symposium therefore reaches a much wider audience nationally and internationally than can possibly attend.

Cold Spring Harbor Laboratory meetings are unique in that organizers are encouraged to select talks from abstracts that are submitted on the basis of scientific merit, ensuring that the latest findings will be presented and that young scientists will have the chance to describe their work. 2017 saw the continuation of many successful annual and biennial meetings as well as the introduction of several new meetings, including STAT Family Proteins, Annexins. The CSHL Genentech Center Conferences on the History of Molecular Biology and Biotechnology this year addressed four decades of RNA-splicing research. The program featured many notable speakers, including John Abelson, Louise Chow, Gideon Dreyfuss, Nouria Hernandez, Adrian Krainer, Thomas Maniatis, James Manley, Melissa Moore, Kiyoshi Nagai, Karla Neugebauer, Richard Roberts, Phillip Sharp, Yigong Shi, David Spector, and Joan Steitz. Partial support for individual meetings is provided by grants from the National Institutes of Health (NIH), National Science Foundation (NSF), foundations, and companies. Core support for the meetings program is provided by the Laboratory's Corporate Sponsor Program.

The Courses program covers a diverse array of topics in molecular biology, neurobiology, structural studies, and bioinformatics. Instructors update their course curricula annually, invite new speakers who bring a fresh perspective, and introduce new techniques and experimental approaches based on student feedback and progress in the field. New techniques—for example, genome editing using tools such as CRISPR or super-resolution microscopy—are introduced as methodologies develop and evolve. We strongly encourage each course to include the latest technical and conceptual developments in their respective fields. Instructors, course lecturers, and course assistants come from universities, medical schools, research institutes, and companies around the world to teach at Cold Spring Harbor. Their excellence and dedication make the course program work uniquely well. We would especially like to thank Drs. Chase Beisel, Luciano Di Croce, Gary Gilliland, Eyal Gottlieb, Andrew Huberman, Rafael Irizarry, Karla Kaun, Chi-Hon Lee, Vincent Noireaux, Stefan Pulver, Howard Salis, and Ali Shilatifard, whose exemplary teaching

and leadership of their respective courses have benefitted so many young scientists. We should especially single out for thanks Drs. Gilliland and Shilatifard, who respectively taught the X-ray crystallography and eukaryotic gene expression courses for many more years than the norm. Students include advanced graduate students, postdoctoral trainees, and principal investigators and senior scientists from around the world.

Grants from a variety of sources support the courses. The core support grants provided through the Helmsley Charitable Trust and Howard Hughes Medical Institute are critical to our course program. The courses are further supported by multiple awards from the NIH and the NSF, and additional support for individual courses is provided by various foundations and other sources. The courses also depend on equipment and reagents that are loaned or donated by a large number of companies—partnerships that are invaluable in ensuring that the courses offer training in the latest technologies (a list of these companies is provided after the Meetings section).

Now in its eighth year of operation, the Cold Spring Harbor Asia (CSHA) program, under the direction of Dr. Maoyen Chi, is headquartered at the Suzhou Dushu Lake Conference Center in the Suzhou Innovation Park high-technology suburb. Seventeen scientific conferences were held in Suzhou and one summer school in Shanghai. CSHA's scientific program is designed for scientists from the Asia/Pacific region, who make up more than 75% of attendance, and include symposia and meetings, training workshops, and occasional Banbury-style discussion meetings. This program is described in more detail in a separate Annual Report.

Special events included the first Double Helix Day event on Discovering Our Ancestry through DNA in February, several bioentrepreneur networking events, and numerous local area one- and two-day retreats, including several affiliated with the Feinstein Institute and Northwell Health. Although distinct from our regular academic program, these events attract significant numbers of leaders and individuals associated with biomedicine and biobusiness from the tristate area and beyond.

The Meetings & Courses program staff comprises a diverse team of talented professionals who handle the complexities of database design, programming, web and multimedia design, educational grants management, marketing and recruitment, conference and course administration, audiovisual and digital design services, and other activities. We said goodbye to Heather Johnson and Michele Navarro, the latter after many years of devoted service to CSHL. We also welcomed several new staff in 2017, including Cathleen Carr, Betty Cavallaro, and Erica Stang, who are already bringing a high level of professionalism to their positions.

David Stewart
Executive Director,
Meetings and Courses Program/
President, Cold Spring Harbor Asia

Terri Grodzicker
Dean of Academic Affairs

COLD SPRING HARBOR ASIA SUMMARY OF CONFERENCES

Dates	*Title*	*Organizers*
April 10–14	Lipid Metabolism and Metabolic Disorders	Peng Li, Peter Tontonoz, Tobias Walther, (Hongyuan) Rob Yang
April 17–21	Bacterial Infection and Host Defense	Kenya Honda, Samuel Miller, Craig Roy, Feng Shao, Jörg Vogel
April 24–28	Cilia and Centrosomes	Monica Bettencourt-Dias, Gert Jansen, Guangshuo Ou, Meng-Fu Bryan Tsou
May 8–12	Francis Crick Symposium: Transforming Neurosciences: Questions and Experiments	Hailan Hu, Edvard Moser, John O'Keefe, Hee-Sup Shin, Alcino Silva
May 15–19	Membrane Proteins: Structure and Function	Martin Caffrey, Nancy Carrasco, Tian-Le Xu, Ming Zhou
May 22–26	Plant Cell and Development Biology	Niko Geldner, Tetsuya Higashiyama, Bo Liu, Yongbiao Xue, Zhenbiao Yang
June 26–30	Primate Neuroscience: Perception, Cognition and Disease Models	Tadashi Isa, Anthony Movshon, Zilong Qiu, Xiaoqin Wang
September 4–8	Microbiota, Metagenomics, and Health	Eran Elinav, Andrew Goodman, Kenya Honda, Zhihua Liu, Andrew Macpherson
September 18–22	Precision Cancer Biology: From Targeted to Immune Therapies	Frederic de Sauvage, Thomas Gajewski, Yutaka Kawakami, Zemin Zhang
October 9–13	Cell Signaling and Metabolism in Development and Disease	Michael Karin, Dianqing Dan Wu, Chenqi Xu, Yingzi Yang
October 16–20	Mitochondria	Paolo Bernardi, Andrew Dillin, Anu Suomalainen-Wartiovaara, Xiaodong Wang
October 22–26	Tumor Immunology and Immunotherapy	Xuetao Cao, Vincenzo Cerundolo, Olivera Finn, Pramod Srivastava
November 6–10	Stem Cells, Aging, and Rejuvenation	Adam Antebi, Jing-dong Han, Brian Kennedy, Seung-Jae Lee, Jan Vijg
November 11–13	Aging and Cancer	Saijuan Chen, Zhu Chen, Samuel Waxman
November 13–17	RNA Modifications and Epitranscriptomics	Michaela Frye, Chaun He, Tsutomu Suzuki, Yungui Yang
December 4–8	Liver, Biology, Diseases, and Cancer	Gen-Sheng Feng, Lijian Hui, Atsushi Miyajima, Lars Zender
December 11–15	Inflammation: Basic Mechanisms and Relevant Diseases	K. Frank Austen, Yasmine Belkaid, Chen Dong

82ND COLD SPRING HARBOR LABORATORY SYMPOSIUM ON QUANTITATIVE BIOLOGY

Chromosome Segregation and Structure

May 31–June 5 282 Participants

ARRANGED BY **Terri Grodzicker, David Stewart, and Bruce Stillman**, Cold Spring Harbor Laboratory

K. Zaret, T. Misteli

The 82nd Cold Spring Harbor Symposium focused on Chromosome Segregation and Structure and addressed the enormous progress in our understanding of segregation, dynamics, and stability of chromosomes—a topic not addressed at a CSHL Symposium for many years. Topics addressed at the Symposium included meiosis; mitosis; chromosome segregation; centrosomes and centrioles, ploidy, chromosome segregation errors, and disease; asymmetric cell division; nuclear architecture; chromosome structure and condensation; sister chromatid cohesion; genome stability; and germ cells. The Symposium attracted more than 275 participants and provided an extraordinary 5-day synthesis of current understanding in the field. Opening night talks setting the scene for later sessions included Steven Henikoff on the inner kinetochore complex, Angelika Amon on cell nonautonomous regulation of chromosome segregation, Scott Keeney on DNA dynamics during meiosis, and Mónica Bettencourt-Dias on centrosome biogenesis. David Page delivered a masterly Dorcas Cummings Lecture on "Sex and Disease: Do Males and Females Read Their Genomes Differently?" for the Laboratory's friends and neighbors. Interviews with leading scientists captured during the Symposium provide a snapshot of the state of current research and are available on the CSHL Leading Strand channel (https://www.youtube.com/user/LeadingStrand).

W. Earnshaw, F. Uhlmann

C. Montagna, G. Andriani

M. Falk

D. Page

B. Tewelde

PROGRAM

Introduction
B. Stillman, *Cold Spring Harbor Laboratory*

Meiosis
Chairperson: S. Gasser, *Friedrich Miescher Institute for Biomedical Research, Basel, Switzerland*

Germ Cells, Imprinting, Gene Dosage, and Regulation
Chairperson: K. Nasmyth, *University of Oxford, United Kingdom*

Centromeres and Kinetochores
Chairperson: Y. Yamashita, *HHMI/University of Michigan, Ann Arbor*

Chromosome Biology and Genome Architecture
Chairperson: M. Jasin, *Memorial Sloan Kettering Cancer Center, New York, New York*

Mitosis and Cell Cycle
Chairperson: G. Almouzni, *Institut Curie/Research Centre, Paris, France*

Chromatin, Histones, and Instability
Chairperson: L. Mirny, *Massachusetts Institute of Technology, Cambridge*

Dorcas Cummings Lecture
Chairperson: D. Page, *HHMI/Whitehead Institute, Massachusetts Institute of Technology, Cambridge*

Organizing Chromosomes during Segregation
Chairperson: S. Cantor, *University of Massachusetts Medical School, Worcester*

Heterochromatin, Errors, and Damage
Chairperson: T. Stearns, *Stanford University, California*

Summary
D. Pellman, *Dana-Farber Cancer Institute, Boston, Massachusetts*

K. Nasmyth, B. Stillman

M. Hetzer, D. Cleveland

MEETINGS

Systems Biology: Global Regulation of Gene Expression

February 26–March 2 223 Participants

ARRANGED BY **Barak Cohen,** Washington University School of Medicine in St. Louis, Missouri
Christina Leslie, Memorial Sloan Kettering Cancer Center, New York, New York
Bas van Steensel, Netherlands Cancer Institute, Amsterdam, Netherlands
Julia Zeitlinger, Stowers Institute for Medical Research, Kansas City, Missouri

Systems Biology is a discipline in which investigators aim to understand the properties of cellular networks by using high-throughput, technology-intensive, and computational approaches. One of the most actively researched areas in systems biology is the global regulation of gene expression, which coordinates complex metabolic and developmental programs in cells, tissues, and organisms. This 11th Systems Biology conference captured the continuing rapid progress in this exciting field. A total of 223 scientists attended the four-day meeting, which featured 64 talks and 98 poster presentations covering a broad range of topics. The speakers, poster presenters, and other conference attendees comprised a mix of students, postdocs, and PIs at all levels.

C. Leslie, B. Cohen, J. Zeitlinger, B. van Steensel

Two keynote speeches were exciting highlights of the meeting. Dr. Patrick Cramer presented important new results documenting progress on assembling a combined crystal structure of many of the proteins in Mediator and how they interact with RNA Pol I. This structure provides an important foundation for understanding the mechanisms by which enhancer activation leads to transcription from the promoter. This fundamental question is also investigated with genomics methods, and Dr. Cramer provided an elegant example of this. Using TT-Seq, he reported a relationship between the distance of RNA Pol II from the initiation site and the level of new initiation. Dr. Xiaowei Zhuang gave the second keynote address and described her lab's work on imaging the three-dimensional conformation of looped DNA using STORM microscopy, which essentially confirmed results from genomics approaches such as Hi-C experiments. There was considerable excitement that the resolutions of new imaging technologies are reaching the point where investigators may soon be able to visualize both the protein and nucleic acid components of individual enhancer–promoter connections.

The addresses by Drs. Cramer and Zhuang echoed key emerging themes in our field. First, it is clear that high-throughput global methods are beginning to make contact with more mechanistic experiments and models. This was exemplified in several other talks about polymerase pausing, enhancer–promoter looping, and the partitioning of active and inactive regulatory DNA into different nuclear compartments. The combination of high-throughput methods with new imaging technologies, as well as studies of cell-to-cell variability, promises to provide in vivo mechanistic insights into global patterns of gene regulation. In line with this trend, single-cell methods continue to grow in prominence, and it appears that sometime soon any assay that can be performed across the genome will have a single-cell counterpart. With the increase in experimental technology, new computational methods to analyze these data are needed and played an important role at the meeting. Another major theme was the increasing attention

S. Bekiranov, A. Hansen

C. Fulco, J. Taylor

being paid to the three-dimensional structure of the genome. Several new technologies and computational models were presented specifically to address how genome structure is determined and what its impact might be on gene regulation. The development of quantitative experimental and computational methods will continue to have a foundational role in our field, and continuing development of these methods is likely to keep this meeting exciting and dynamic for several years to come.

This meeting was supported with funds provided by the National Institutes of Health.

PROGRAM

cis-Regulatory Logic
Chairpersons: A. DePace, *Harvard Medical School, Boston, Massachusetts;* J. Wysocka, *Stanford University, California*

Chromatin Biology and 3D Architecture of the Nucleus
Chairpersons: L. Mirny, *Massachusetts Institute of Technology, Cambridge;* A. Raj, *University of Pennsylvania, Philadelphia*

Keynote Speaker
P. Cramer, *Max Planck Institute for Biophysical Chemistry, Göttingen, Germany*

Cellular Decision-Making
Chairpersons: E. Segal, *Weizmann Institute of Science, Rehovot, Israel;* I. Amit, *Weizmann Institute of Science, Rehovot, Israel*

Evolution and Variation
Chairpersons: K. Pollard, *University of California, San Francisco;* T. Lappalainen, *New York Genome Center, Columbia University, New York*

Transcription Factors
Chairpersons: Karen Adelman, *Harvard Medical School, Boston, Massachusetts;* D. Trono, *Ecole Polytechnique Fédérale de Lausanne, Switzerland*

Keynote Speaker
X. Zhuang, *Harvard University, Cambridge, Massachusetts*

RNA and Its Regulation
Chairpersons: M. Zavolan, *University of Basel, Switzerland;* G. Yeo, *University of California, San Diego*

Emerging Technologies
Chairpersons: S. Morris, *Washington University School of Medicine in St. Louis, Missouri;* D. Larson, *National Cancer Institute, National Institutes of Health, Bethesda, Maryland*

Networks, Global Analysis, and Computational Models
Chairpersons: O. Stegle, *EMBL-EBI, Hinxton, United Kingdom;* A. Kundaje, *Stanford University, California*

T. Lappalainen, Y. Shen

Double Helix Day: Discovering Our Ancestry through DNA

March 7 80 Participants

ARRANGED BY **David Stewart and Bruce Stillman,** Cold Spring Harbor Laboratory

This special annual celebration ("Double Helix Day") is intended to coincide with the actual date on which James Watson and Francis Crick discovered the double helix structure of deoxyribonucleic acid (February 28, 1953) in Cambridge, England. Each year, a theme related to DNA science is explored through a series of review-style talks aimed at a broad scientific audience. The theme Discovering Our Ancestry through DNA was selected as the theme of the 2017 celebrations. The program commenced in the early afternoon on Tuesday, March 7, and concluded with an early evening reception and celebratory double helix feast in Blackford Hall.

C. Bustamante, J. Watson

Invited Speakers

C. Ball, *Ancestry, Lehi, Utah:* Population structure of North America: Insights from one million genomes.

C. Bustamante, *Stanford University, California*: Peopling of the Americas.

P. Donnelly, *University of Oxford, Oxford, United Kingdom:* Fine genetic structure of the UK.

Y. Erlich, *New York Genome Center/Columbia, New York*: Crowd genomics.

D. Stewart, C. Ball

Y. Erlich

Systems Biology: Networks

March 14–18 121 Participants

ARRANGED BY **Pascal Falter-Braun,** Helmholtz Zentrum Münich & Ludwig-Maximilians-University, Germany
Suzanne Gaudet, Dana-Farber Cancer Institute, Boston, Massachusetts
Ben Lehner, Center for Genomic Regulation, Barcelona, Spain
Chad Myers, University of Minnesota, Minneapolis

S. Gaudet, C. Myers, B. Lehrer, P. Falter-Braun

In cells and organisms, genomic information is translated into phenotypes by complex and highly dynamic molecular networks formed by proteins, nucleic acids, and small molecules. A systems-level understanding of biological systems, as well as the design of rational biotechnological or pharmaceutical interventions in human, crops, and microbes, hinges on our knowledge of these networks.

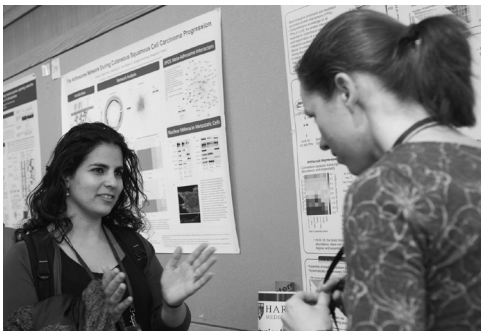
As our knowledge of molecular networks remains largely incomplete, an important goal of biological network science is to experimentally map or computationally infer the wiring of cells. A second major goal is the mechanistic characterization of smaller network modules, translating large-scale network connectivity into molecular mechanisms as a basis for developing quantitative and predictive models. Finally, network science also aims to develop and apply statistical tools to extract insights from known biological networks to identify disease-causing genes and modules, identify targets for intervention, and decipher the fundamental principles that underlie biological systems and their evolution. Two key goals of this meeting remain to bridge these three aspects of network biology and to cover a diversity of biological systems, from humans and model organisms to plants and microbes.

In the 2017 meeting, we continued two successful innovations introduced in 2015. One was an informal "Lunch with a PI" to foster interactions between established PIs and young scientists—about both science and career development. The second was an open discussion that this year addressed the issues of statistics and reproducibility in systems biology. This highly interactive session involved panelists (statisticians, biologists, and a journal editor) and plenty of audience participation. These two elements were again praised by attendees and resulted in continued discussions throughout the meeting.

Although the opening of the scientific program was delayed by a snowstorm, the full program was fit into the remaining available time. There were 15 invited presentations and 25 short talks selected from submitted abstracts, all of them outstanding and many given by postdoc and Ph.D. students. The talks covered a range of concepts—from the importance of dynamics in signaling networks to the development of new methods to map the complete human protein interaction network—and addressed diverse questions from plant and animal development to cancer and other human diseases, highlighting how systems biology brings together people from different fields of biology. These presentations showcased recent advances and also the open questions, exciting opportunities, and technological challenges that remain. A particularly thought-provoking talk from Hiroaki Kitano presented the grand challenge of automating the scientific process using artificial intelligence.

The two keynote addresses were inspiring for Ph.D. students and postdocs and presented a historical perspective on how the field has developed during the past 10–20 years. Marian Walhout opened the meeting with an overview of her lab's efforts to systematically map gene regulatory

F. Barrenas, J. Komorowski, M. Bassik



N. Shemesh, K. Luck

networks and how they are perturbed in disease. Marc Vidal closed the meeting with the take-home message that altered interactions ("edges") is a powerful framework to understand human genetic disease, and he helped lead a community discussion on the future of the field.

This meeting was funded in part by the National Human Genome Research Institute, a branch of the National Institutes of Health.

PROGRAM

Keynote Speaker
M. Walhout, *University of Massachusetts Medical School, Worcester*

Regulatory Networks
Chairperson: C. Myers, *University of Minnesota, Minneapolis*

Interaction Networks I
Chairperson: M. Walhout, *University of Massachusetts Medical School, Worcester*

Network Evolution
Chairperson: B. Lehner, *Centre for Genomic Regulation, Barcelona, Spain*

Lightning Talks
Chairperson: B. Lehner, *Centre for Genomic Regulation, Barcelona, Spain*

Interaction Networks II
Chairperson: T. Ideker, *University of California, San Diego*

Networks and Disease I
Chairperson: F. Roth, *University of Toronto, Canada*

Microbial, Metabolic, and Synthetic Networks
Chairperson: P. Falter-Braun, *Helmholtz Zentrum München & Ludwig-Maximilians-Universität München, Germany*

Panel Discussion: Statistics and Reproducibility in Systems Biology

Signaling Networks
Chairperson: M. Vidal, *Dana-Farber Cancer Institute/ Harvard Medical School, Boston, Massachusetts*

Modeling Network Dynamics
Chairperson: K. Miller-Jensen, *Yale University, New Haven, Connecticut*

Networks and Disease II
Chairperson: B. Andrews, *University of Toronto, Canada*

Network Pharmacology
Chairperson: S. Gaudet, *Dana-Farber Cancer Institute/ Harvard Medical School, Boston, Massachusetts*

Genetic Networks
Chairperson: J. Taipale, *Karolinska Institutet, Stockholm, Sweden*

Closing Keynote Speaker
M. Vidal, *Dana Farber Cancer Institute/Harvard Medical School, Boston, Massachusetts*

B. Lehner, M. Springer

RNA and Oligonucleotide Therapeutics

March 29–April 1 124 Participants

ARRANGED BY **Annemieke Aartsma-Rus,** Leiden University Medical Center, Netherlands
Matthew Levy, Albert Einstein College of Medicine, Bronx, New York
Laura Sepp-Lorenzino, Alnylam Pharmaceuticals, Cambridge, Massachusetts

A. Aartsma-Rus, L. Sepp-Lorenzino

This fifth conference focused on the development of oligonucleotides as drugs, covering both preclinical work and clinical trials. It brought together top scientists from academia and industry interested in a variety of RNA-based therapeutics. The theme "from the bench to the bedside" made it unique among CSHL meeting series. The meeting opened with a keynote by Dr. Anastasia Khvorova on siRNA and delivery. Covered approaches ranged from single-stranded antisense oligonucleotides for splicing modulation, targeted transcript knockdown, aptamers and immunostimulation, siRNAs and antagomirs, to mRNA therapeutics. The applications of these varied technologies included rare diseases, cancer, cardiovascular diseases, and bacterial and viral infections. The all-important issue of RNA therapeutics delivery by a variety of methods was also covered during talks and a panel discussion.

The participants came from 11 companies and from universities and research institutions in the United States and abroad. The seven scientific sessions featured 37 platform talks, 20 posters, and a panel discussion. Animated and insightful exchanges during the sessions continued throughout. A number of participants expressed interest in attending this conference next year.

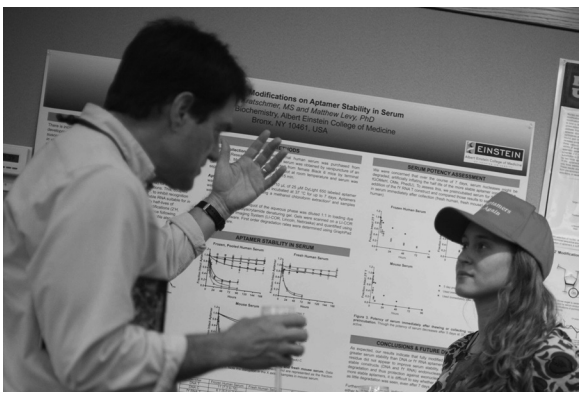
This meeting was funded by Arcturus Therapeutics, Alnylam Pharmaceuticals, Moderna Therapeutics, and RaNA Therapeutics.

PROGRAM

Keynote Speaker
A. Khvorova, *University of Massachusetts Medical School, Worcester*

Double-Stranded RNA
Chairpersons: L. Sepp-Lorenzino, *Alnylam Pharmaceuticals Inc., Cambridge, Massachusetts;* D. Lewis, *Madison, Wisconsin*

A. Watkins, D. Dreyfus



S. Dowdy, C. Kratschmer

Aptamers
Chairpersons: M. Levy, *Albert Einstein College of Medicine, Bronx, New York;* D. Burke, *University of Missouri, Columbia*

Short Talks and Round-Table Discussion
Chairperson: A. Aartsma-Rus, *Leiden University Medical Center, Netherlands*

Round Table Discussion: Delivery Opportunities and Challenges

Single-Stranded Oligonucleotides
Chairpersons: A. Krainer, *Cold Spring Harbor Laboratory;* A. Aartsma-Rus, *Leiden University Medical Center, Netherlands*

Oligonucleotide Therapeutics Development
Chairpersons: A. Levin, *Avidity Biosciences LLC, La Jolla, California;* L. Sepp-Lorenzino, *Alnylam Pharmaceuticals Inc., Cambridge, Massachusetts*

A. Naar, I. Jarchum

mRNA Therapeutics
Chairpersons: M. Moore, *Moderna Therapeutics, Cambridge, Massachusetts;* M. Fotin-Mleczek, *CureVac AG, Tübingen, Germany*

Wiring the Brain

April 4–8 125 Participants

ARRANGED BY **Danielle Bassett,** University of Pennsylvania, Philadelphia
Joshua Gordon, National Institute of Mental Health, Bethesda, Maryland
Josh Huang, Cold Spring Harbor Laboratory
Kevin Mitchell, Trinity College Dublin, Ireland

J. Huang, J. Gordon, D. Bassett, K. Mitchell

This third conference followed those in 2013 and 2015 and two previous meetings held in Ireland in 2009 and 2011. The main goal of this meeting is to bring together researchers from diverse fields to explore how brain connectivity is established, how genetic variation can affect these processes, how circuit and network function are affected by defects in neural development, and how this can lead to psychiatric and neurological disease. The program was designed with a particular focus on neurodevelopmental disorders and covered a wide range of areas relevant to this theme. The session topics are listed below in the Program.

We had 20 invited talks from world leaders in diverse fields and an additional 18 talks selected from submitted abstracts. Not only did these talks highlight cutting-edge research in specific fields, notably many of them were integrating across levels of analysis, using new technologies to transition smoothly from cells to circuits to systems, from development to function or dysfunction, and between animals and humans.

Two stimulating keynote lectures set an excellent tone for the meeting. The first, by Karl Deisseroth, described the development of optogenetic tools that are revolutionizing neuroscience. These tools provide the means to selectively activate or inhibit genetically defined subsets of neurons with millisecond precision, in awake, behaving animals. Since their development by Deisseroth and colleagues in 2009, these tools have been used to dissect the circuit mechanisms underlying an incredible array of brain functions and behaviors.

The second lecture was by philosopher Patricia Churchland, who discussed the evolved, neural basis of moral behaviors in animals and humans. Churchland has been a leader in incorporating detailed neuroscientific knowledge into a philosophical framework for understanding human behavior, and this talk exemplified how both science and philosophy are required to address these kinds of big questions. Together, these two lectures illustrated the necessity and the possibility of linking analyses across multiple levels, from molecular and cellular, to circuits, systems, and behavior, including human behaviors and the emergence of culture.

We also had 86 posters, which were similarly outstanding in the quality of research presented, again across very diverse fields. Poster sessions were very well attended, lively, and stimulating.

The meeting was characterized by a very open attitude and wide-ranging discussions centered on big questions, rather than focused on technical details. As such, it provided a unique forum for the integration of research from developmental neurobiology, psychiatric genetics, molecular, cellular and systems neuroscience, cognitive science, and psychology. These kinds of interactions will help build an integrative explanatory framework in neuroscience, especially in our understanding of the nature and causes of brain disorders.

This meeting was funded in part by Lundbeck.

D. Margolis, S. Sorensen

T. Zador, K. Deisseroth

PROGRAM

Development: Early Events
Chairperson: K. Mitchell, *Trinity College Dublin, Ireland*

(Human) Brain Development
Chairperson: J. Huang, *Cold Spring Harbor Laboratory*

Neurodevelopmental Disorders
Chairperson: K. Mitchell, *Trinity College Dublin, Ireland*

Keynote Speaker
K. Deisseroth, *Stanford University*

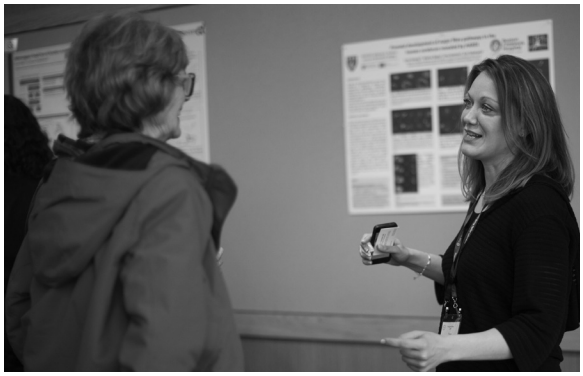
Synapse Formation and Plasticity
Chairperson: J. Huang, *Cold Spring Harbor Laboratory*

Sexual Differentiation
Chairperson: D. Bassett, *University of Pennsylvania, Philadelphia*

Microcircuits
Chairperson: D. Bassett, *University of Pennsylvania, Philadelphia*

Systems: Function and Dysfunction
Chairperson: J. Gordon, *National Institute of Mental Health, Bethesda, Maryland*

Keynote Speaker
P. Churchland, *University of California, San Diego*



P. Churchland, L. Vasung

Cellular Dynamics and Models

April 11–14 103 Participants

ARRANGED BY **J. Faeder,** University of Pittsburgh, Pennsylvania
E. Munro, University of Chicago, Illinois

E. Munro, J. Faeder

The 2017 meeting maintained an emphasis on quantitative approaches to cell biology. The scientific program comprised two keynote talks, themed sessions combining talks from invited speakers with shorter contributed talks, a popular sequence of afternoon tutorials on computational approaches, and an afternoon poster session.

On Tuesday night, Mike Rust and Jan Skotheim led an exciting session highlighting quantitative analysis of oscillatory dynamics in single cells.

On Wednesday morning, two sessions highlighted biophysical approaches to subcellular organization and dynamics with talks on curvature sensing by membrane proteins (Patricia Bassereau), single-molecule analysis of signaling dynamics (Jay Groves), directional force sensing by adhesion proteins (Alex Dunn), and the polymer physics of homology search during double-stranded break repair in yeast (Jane Kondev). This was followed by a very lively afternoon poster session.

In Wednesday evening's keynote talk, Tom Pollard discussed both historical and more recent efforts to understand the dynamics of cytokinesis and endocytosis in fission yeast, through the integration of biochemistry, live-cell measurements, and mathematical modeling. Dyche Mullins continued the cytoskeletal theme by describing his recent efforts to understand physical and chemical principles that allow branched actin networks to act as spatially distributed motors.

Thursday morning's first session focused on signaling dynamics. Bill Hlavacek described a systems-level analysis of autophagy using integrated computational and experimental approaches.

G. Fjeld, J. Kim

H. Kugler, J.G.T. Zanudo

A. Dunn, T. Pollard

D. Striegel, C. Miller

Additional talks emphasized how rapid advances in quantitative proteomics, CRISPR-Cas9 gene editing, and single-cell and single-molecule imaging are deepening our understanding of cell signaling.

In a second session, Boris Shraiman discussed how local distributions of Myosin II activity along cell–cell junctions drive tissue-level flows during early development. Lani Wu then described the self-organizing principles that allow developing neurons to map spatial information from *Drosophila* ommatidia into the central nervous system.

The popular set of tutorial sessions was followed by a second keynote lecture by Rong Li on describing experimental and theoretical analysis of how chromosome instability can mediate adaptation to stress, with implications for cancer therapy.

On Friday morning, Sharad Ramanathan described statistical approaches to inferring cell-state transitions during early mammalian development from single-cell gene expression. Graham Johnson and colleagues gave an exciting overview of ongoing work at the new Allen Institute for Cell Science, fusing cell engineering, advanced light microscopy, and computer vision.

In summary, this year's meeting was an unqualified success. The attendance was up again. The quality of talks was exceedingly high. Individual sessions were uniformly lively, and the discussions over dinner and late-night drinks were equally intense. It is clear that this meeting continues to fill an important niche in the rapidly growing area of quantitative biology, with a unique ability to draw and excite representatives from many areas of biology, representing a broad array of approaches to quantitative analysis of biological dynamics.

In 2019, James Faeder and Ed Munro will be joined by Susan Rafelski of the Allen Institute for Cell Science and Dyche Mullins of the University of California, San Francisco. We will follow the same overall format, but because of popular demand, (1) we propose to move the poster sessions into the evenings to allow extended discussions and to replace the final evening banquet with a second poster session to allow more time for poster presenters to show their work and see the work of others, and (2) we also propose to diversify the afternoon tutorials and replace longer tutorials with one or more regular sessions that highlight exemplary uses of computational, imaging, and computer vision approaches. The organizers look forward to 2019!

PROGRAM

Single-Cell Dynamics
Chairpersons: J. Skotheim, *Stanford University, California*; M. Rust, *University of Chicago, Illinois*

Membrane Dynamics
Chairpersons: P. Bassereau, *Institut Curie, Paris, France;* J. Groves, *University of California, Berkeley*

Subcellular Dynamics
Chairpersons: A. Dunn, *Stanford University, California*; J. Kondev, *Brandeis University, Waltham, Massachusetts*

ADVANCED TUTORIAL SESSIONS

Keynote Speaker
T.D. Pollard, *Yale University, New Haven, Connecticut*

Self Organization of the Cytoskeleton
Chairperson: D. Mullins, *University of California, San Francisco*

Signaling Dynamics
Chairperson: W. Hlavacek, *Los Alamos National Laboratory, New Mexico*

Tissue Dynamics
Chairperson: B. Shraiman, *University of California, Santa Barbara*

Keynote Speaker
R. Li, *Johns Hopkins University, Baltimore, Maryland*

Dynamics of Cell Fate Determination
Chairperson: S. Ramanathan, *Harvard University, Cambridge, Massachusetts*

Computational Image Analysis
Chairpersons: G. Johnson, *Allen Institute for Cell Science, Seattle, Washington;* L. Wu, *University of California, San Francisco*

The Ubiquitin Family

April 18–22 171 Participants

ARRANGED BY **Ronald Hay,** University of Dundee, United Kingdom
Michael Rape, University of California, Berkeley
Cynthia Wolberger, Johns Hopkins University School of Medicine, Baltimore, Maryland

R. Hay, C. Wolberger, M. Rape

Following the successful inauguration of this series in 2003, this seventh meeting focused on ubiquitin and a group of structurally related "ubiquitin-like" proteins and their roles in regulation of various cellular processes. Major questions in the field include how specificity in ubiquitin and ubiquitin-like protein conjugation reactions is maintained, the molecular mechanisms by which specific ubiquitin chain linkages are recognized and regulate distinctive pathways, the roles of the ubiquitin system in physiology and disease, and opportunities and mechanisms for therapeutically targeting the ubiquitin system. These questions are being elegantly addressed using structural biology, sophisticated kinetic studies, systematic library screening, and quantitative proteomics technologies by an increasing number of investigators in the field. In addition, important advances continue to be made in understanding how ubiquitin and its family members contribute to the operation of diverse cellular pathways and how these functions are perturbed in diseases ranging from viral infections to neurodegenerative diseases and cancers. Exciting progress is also being made in the development of small-molecule inhibitors of a range of enzymes involved in ubiquitin and ubiquitin-like protein pathways.

The meeting was capped by two exciting keynote lectures from Christopher Lima, a Howard Hughes Medical Institute (HHMI) investigator and structural biologist who has produced seminal work in deciphering the molecular mechanism of ubiquitylation enzymes, and Anne Bertolotti, who has published groundbreaking studies on proteasome maturation and its function in neurodegenerative diseases. The meeting attracted scientists who engaged in lively discussions concerning the roles of ubiquitin in protein turnover, quality control, or DNA-damage repair. Quality control was discussed

R. Beale, Y. Merbl, R. Benyair

N. Zheng, Y. Kudo

from a point of protein misfolding, which often accounts for neurodegenerative diseases, but also in terms of co-translational quality control mechanisms that allow a cell to rescue a stalled ribosome. In this context, new ubiquitylation signals, such as branched ubiquitin chains or ubiquitin conjugates containing K6-linkages, were introduced for the first time into the ubiquitin field. These findings underscored the complexity of this field, as well as the fast pace of new discoveries in this important area of biology. The session focused on DNA repair underscored the role of ubiquitin in the nucleus, as well as its important functions in preventing tumorigenesis; indeed, this reflects the current understanding that mutations in ubiquitylation enzymes account for a large number of distinct cancer types. Analysis of proteasome structure and function highlighted the complexity of this machine—that is, the many distinct yet specific ubiquitylation receptors as well as the major conformational changes it undergoes during a catalytic cycle. These studies are important because proteasome inhibitors have become a staple in treating myeloma or mantle cell leukemia. Another session focused on ubiquitin-like proteins, a family of cellular factors that resemble ubiquitin but are attached to substrates by less complex enzymatic cascades. Such "UBLs" have therefore often been used to discover principles of ubiquitin-dependent signaling, which was again the case during this meeting. Reflecting an increasing appreciation in the ubiquitin field and beyond that dynamics of ubiquitylation is often central to signaling, this meeting also discussed the increasing importance of deubiquitylating enzymes in ubiquitin-dependent signaling pathways. Both ubiquitylation enzymes and deubiquitylases have recently been shown to be targets of small molecules used in chemotherapy, and, accordingly, the role of such molecules or new chemical strategies to modulate the activity of cellular ubiquitylation pathways was a focus of this meeting's discussion. Finally, a recent important development in the ubiquitin field has been the identification of important ubiquitylation events during cell fate specification and differentiation, an area that was also discussed in depth during this meeting. It should be noted that several highlights of this meeting were provided by student or postdoc talks that greatly contributed to its cutting-edge character. In conclusion, this is a major meeting where new and unpublished work in the essential—and increasingly druggable—ubiquitylation pathway is being discussed, and it therefore has an essential role for communities interested in understanding key signal transduction pathways.

K. Majorek, M. Nakasone, A. Patel

PROGRAM

Mechanisms of Ubiquitination
Chairpersons: B. Schulman, *HHMI/St. Jude Children's Research Hospital, Memphis, Tennessee;* N. Zheng, *HHMI/University of Washington, Seattle*

Keynote Speaker
C.D. Lima, *HHMI/Sloan Kettering Institute, New York, New York*

Protein Quality Control
Chairpersons: J. Frydman, *Stanford University, California;* R. Kopito, *Stanford University, California*

Ubiquitin/UBLs in the DNA Damage Response
Chairpersons: N. Mailand, *University of Copenhagen, Denmark;* R. Greenberg, *University of Pennsylvania, Philadelphia*

Proteasome
Chairpersons: A. Matouschek, *University of Texas, Austin;* K. Walters, *National Cancer Institute, Frederick, Maryland*

UBLs
Chairpersons: A. Dejean, *Institut Pasteur, Paris, France;* F. Melchior, *Heidelberg University, Germany*

Ubiquitin Removal and Recognition
Chairpersons: D. Komander, *MRC Laboratory of Molecular Biology, Cambridge, Massachusetts;* N. Thoma, *Friedrich Miescher Institute for Biomedical Studies, Basel, Switzerland*

A. Bertolotti, J.W. Harper

Cellular Roles of Ubiquitin
Chairpersons: I. Dikic, *Goethe University, Frankfurt/Main, Germany;* M. Pagano, *HHMI/New York University Langone Medical Center, New York*

Keynote Speaker
A. Bertolotti, *MRC Laboratory of Molecular Biology, Cambridge, United Kingdom*

Ubiquitin in Pathogenesis
Chairpersons: K. Rittinger, *The Francis Crick Institute, London, United Kingdom;* F. Randow, *MRC Laboratory of Molecular Biology, Cambridge, United Kingdom*

Fundamental Immunology and Its Therapeutic Potential

April 25–29 207 Participants

ARRANGED BY **Eric Pamer,** Memorial Sloan Kettering Cancer Center, New York, New York
Fiona Powrie, University of Oxford, United Kingdom
Stephen Smale, University of California School of Medicine, Los Angeles

F. Powrie, S. Smale

This second meeting on fundamental immunology was held on cool but sunny spring days. The meeting was inspired by, and continued to build on, the remarkable success of cancer immunotherapy, in which substantial fractions of patients with cancers that previously had average survival lengths measured in months now have durable responses with low rates of relapse. The success of cancer immunotherapy has boosted the morale of the entire immunology field by showing that fundamental advances in our understanding of immune regulation can lead to profound lifesaving therapies. However, despite the successes, much remains to be learned, as some types of cancer fail to respond to current immunotherapies, and, for cancers with relatively high response rates (e.g., melanoma and lung cancer), strategies for further improvement are desperately needed and are well within reach. Successes have also been emerging in the development of therapies of other diseases linked to the immune system, including infectious diseases, autoimmunity, and a variety of inflammatory diseases. It is anticipated that future advances will require continued interaction between translational researchers studying a diverse range of diseases and basic researchers focused on fundamental aspects of immune regulation. The meeting included eight sessions with oral presentations by invited speakers and speakers selected from submitted abstracts and two engaging poster sessions. The first session focused on fundamental immunology, including the discovery of novel immunoregulatory mechanisms and technological advances in single-cell gene expression profiling and high-resolution imaging. Next came a series of presentations on antimicrobial immunity and innovative vaccination strategies, including presentations by Pamela Bjorkman

P. Tong, H. Algood

L. Messina, W. Koff

M. Crane, I. Odell

P. Savage, A. Chawla

(Caltech) and Michel Nussenzweig (Rockefeller University), who described the surprising finding that individuals who developed high titers of neutralizing antibody to Zika virus exhibited evidence of previous exposure to dengue 1 virus. The sessions on antitumor immunity focused heavily on efforts to better define the mechanisms by which checkpoint blockade therapies and chimeric antigen receptor therapies lead to reactivation of the immune system in the tumor microenvironment. An entire session was devoted to the study of T-regulatory cells, which are thought to have critical roles in immune suppression in tumors and infectious diseases, yet have considerable therapeutic potential for the control of autoimmune diseases. Studies of the microbiome in healthy and disease settings represented a final focal point for the conference, because of emerging evidence that immune homeostasis is strongly influenced by commensal microorganisms and that changes in the microbiota can have both beneficial and harmful effects. Cold Spring Harbor Laboratory continued to be an ideal venue for promoting interactions between leading fundamental and translational immunologists because of the intimate setting, in which all participants dined together at Blackford Hall and met for extended discussions after the evening sessions in the only accessible bar.

PROGRAM

Fundamental Immunology
Chairperson: H. Singh, *Cincinnati Children's Hospital Medical Center, Ohio*

Vaccines and Antiviral Immunity
Chairperson: P. Bjorkman, *HHMI/California Institute of Technology, Pasadena*

Antitumor Immunity
Chairperson: F. Powrie, *University of Oxford, United Kingdom*

T-Regulatory Cells
Chairperson: A. Rudensky, *HHMI/Memorial Sloan Kettering Cancer Center, New York, New York*

Tumor Microenvironment and ILCs
Chairperson: M. Li, *Memorial Sloan Kettering Cancer Center, New York, New York*

Metabolism and Microbiota
Chairperson: Y. Belkaid, *NIAID, National Institutes of Health, Bethesda, Maryland*

Microbiota and NK Cells
Chairperson: E. Pamer, *Memorial Sloan Kettering Cancer Center, New York, New York*

T-Cell Differentiation and Antimicrobial Immunity
Chairperson: S. Kaech, *Yale University School of Medicine, New Haven, Connecticut*

Telomeres and Telomerase

May 2–6 279 Participants

ARRANGED BY **Steven Artandi,** Stanford University, California
Julia Cooper, National Cancer Institute, Bethesda, Maryland
Roger Reddel, Children's Medical Research Institute, Westmead, New South Wales, Australia

This 10th conference in the series, which has been held every two years from 1999 onward, consisted of eight sessions of talks and two poster sessions. As for every one of these meetings, the format was to invite two chairs per session, who were a mix of established scientists in the field and younger scientists who had already made their mark as independent investigators. Session chairs were given the choice between giving a 12-minute presentation themselves or having a member of their lab give a talk. The remainder of the talks (also 12 min) were chosen from submitted abstracts, allowing as many presentations as possible. These presentations were primarily by graduate students and postdoctoral fellows. There were nearly 280 participants, a large fraction of whom presented the 130 posters and 76 talks.

S. Artandi, J. Cooper, R. Reddel

The talks and posters covered diverse aspects of telomere and telomerase biology, including regulation of telomerase expression and activity, telomerase biogenesis and structure, telomere replication, mechanisms of ALT, mechanisms of telomere protection, telomere protein functions at telomeres and throughout the genome, telomere shortening and mechanisms of senescence and aging, and the role of telomeres in human health and disease.

The scientific content was very high throughout the conference in both the talks and the posters. Most of the presented data were unpublished and extensively discussed in an open fashion. Formal and informal discussions were lively and informative. The conference was judged to be highly successful based on verbal and e-mail communications to the organizers. There is strong enthusiasm for another meeting on the same topic in 2019.

Wine and Cheese, Nicholls Biondi

PROGRAM

Regulation of Telomerase Expression and Activity
Chairpersons: C. Greider, *Johns Hopkins University, Baltimore, Maryland;* D. Hockemeyer, *University of California, Berkeley*

Telomerase Biogenesis and Structure
Chairpersons: G. Raffa, *Sapienza University of Rome, Italy;* J. Feigon, *University of California, Los Angeles*

Telomere Replication
Chairpersons: Z. Songyang, *Baylor College of Medicine, Houston, Texas/Sun Yat Sen University, Guangzhou, China;* T. Cech, *HHMI/University of Colorado, Boulder*

Mechanisms of ALT
Chairpersons: H. Pickett, *University of Sydney, Westmead, New South Wales, Australia*; R. Greenberg, *University of Pennsylvania, Philadelphia*

Mechanisms of Telomere Protection
Chairpersons: A. Sfeir, *New York University School of Medicine;* F. Ishikawa, *Kyoto University, Japan*

Telomere Protein Functions at Telomeres and Genome-Wide
Chairpersons: V. Geli, *CNRS-INSERM, Marseille, France;* K. Friedman, *Vanderbilt University, Nashville, Tennessee*

T. Cech, M. Armanios

Telomere Shortening and Mechanisms of Senescence
Chairpersons: T. de Lange, *The Rockefeller University, New York;* W. Wright, *University of Texas Southwestern Medical School, Dallas*

Telomeres and Human Health and Disease
Chairpersons: A. London-Vallejo, *Institut Curie, Paris, France;* M. Armanios, *Johns Hopkins University School of Medicine, Baltimore, Maryland*

K. Karimian, C. Greider

K. Collins, D. Shippen

The Biology of Genomes

May 9–13 533 Participants

ARRANGED BY **Michel Georges,** University of Liege, Belgium
Matthew Hurles, Wellcome Trust Sanger Institute, Hinxton, United Kingdom
Dana Pe'er, Columbia University, New York, New York
Jonathan Pritchard, Stanford University, California

M. Georges, D. Pe'er, M. Hurles, J. Pritchard

This annual meeting marked the 30th annual gathering of genome scientists at Cold Spring Harbor Laboratory. Investigators from around the world attended the meeting, with more than 350 abstracts presented describing a broad array of topics relating to the functional analysis, comparative characterization, and interpretation of genomes from diverse organisms. The scope and applicability of genome science continues to grow; even after almost three decades of this meeting, the talks were fresh and there was fierce competition for abstracts. It featured talks on a wide variety of genomics topics such as translational genomics and genetics, population genomics, genetics of complex traits, cancer and medical genomics, complex traits and microbiome, evolutionary and nonhuman genomics, functional genomics, and computational genomics. Session chairs ensured a reasonable balance of genders, and there was a strong focus on younger graduate students and postdoc presentations.

The talks ranged from ecological/evolutionary biology (e.g., Karine van Doninck on the genomics of rotifers) through to cancer genomics (Elaine Mardis on breast cancer genomics) and functional genomics (Jay Shendure). Talks included engaging examples of genomics expanding in scope: for example, Barbara Treutlein on genomic applications to in vitro organoids. The impact of genomics in translation studies was exemplified by a talk from Matthew Garnett. Several talks focused on genomic approaches to the microbiome (e.g., Moran Yassour and Ran Blekhman). Three poster sessions allowed for comprehensive discussions of abstracts that did not make the talks; they were well attended.

These topics gave evidence of DNA sequence variation and its role in molecular evolution, population genetics and complex diseases, comparative genomics, large-scale studies of gene and

A. Valouev, D. Schwartz

S. Pääbo, K. Frazer

H. Ollila, J. Humphrey

R. Hardison, R. Gibbs

protein expression, and genomic approaches to ecological systems. Both technologies and applications were emphasized.

All sessions were well attended, stretching the capacity of the CSHL facilities. The keynote presentations were from Andrew Clark, whose topic was Junk Evolution, and Aviv Regev, who discussed reconstructing intra- and intercellular circuits with single-cell genomics.

The ELSI (Ethical, Legal, and Social Implications) panel was moderated by Dave Kaufman, National Human Genome Research Institute, National Institutes of Health, and the area of discussion was "What's in a Name? Diversity and the Future of Genomic Research." Panelists included Eimar Kenny, Icahn School of Medicine at Mt. Sinai; Sandra Lee, Stanford University; Alondra Nelson, Columbia University; and Aliya Saperstein, Stanford University.

This meeting was funded in part by the National Human Genome Research Institute, a branch of the National Institutes of Health; Illumina; Nanopore Technologies; and Swift Biosciences.

PROGRAM

Cancer and Medical Genomics
Chairpersons: E. Mardis, *Nationwide Children's Hospital, Columbus, Ohio;* C. Curtis, *Stanford University, California*

Complex Traits and Microbiome
Chairpersons: B. Neale, *Massachusetts General Hospital, Boston;* M. Yassour, *Broad Institute of MIT/Harvard, Cambridge, Massachusetts*

Evolutionary and Non-Human Genomics
Chairpersons: K. van Doninck, *Université de Namur, Belgium;* N. Phadnis, *University of Utah, Salt Lake City*

Functional Genomics
Chairpersons: H. Chang, *Stanford University School of Medicine, California;* B. Treutlein, *Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany*

B. Engelhardt, E. Lander

ELSI Panel Discussion

What's in a Name? Diversity and the Future of Genomic Research
D. Kaufman, *NIH/National Human Genome Research Institute, Bethesda, Maryland*

Translational Genetics and Genomics
Chairpersons: M. Garnett, *Wellcome Trust Sanger Institute, Hinxton, United Kingdom;* J. Shendure, *University of Washington, Seattle*

Population Genomics
Chairpersons: M. Schumer, *Harvard University, Boston, Massachusetts/Columbia University, New York, New York;* S. Sunyaev, *Brigham & Women's Hospital/Harvard Medical School, Boston, Massachusetts*

Guest Speakers
Chairpersons: A. Clark, *Cornell University, Ithaca, New York;* A. Regev, *Broad Institute of MIT/Harvard, Cambridge, Massachusetts*

Computational Genomics
Chairpersons: A. Kundaje, *Stanford University, California;* A. Battle, *Johns Hopkins University, Baltimore, Maryland*

Mechanisms of Metabolic Signaling

May 16–20 153 Participants

ARRANGED BY **Mitchell Lazar,** University of Pennsylvania, Perelman School of Medicine, Philadelphia
Susanne Mandrup, University of Southern Denmark, Odense
Jared Rutter, HHMI/University of Utah School of Medicine, Salt Lake City

J. Rutter, S. Mandrup, M. Lazar

The main goal of this meeting was to bring together researchers from diverse fields to explore how principles of cellular metabolism manifest in different cell types; how metabolic regulation underlies the functions of specialized tissues; and how these differences impact both normal physiology and diseases such as diabetes and cancer.

The 24 invited speakers were leaders in the various aspects of metabolic research from all over the world, and the majority had not been part of the 2015 program. The meeting opened with an inspiring keynote address by Lasker Awardee Bill Kaelin, which set a superb tone for the entire meeting. A total of 10 sessions followed, as well as a second keynote talk by Steve McKnight that added to the gravitas of the program. Eight sessions featured oral presentations, all highlighting unpublished research and focused on key areas in the field of metabolism. Emphasis included, but was not limited to, genomic and epigenomic mechanisms, signalling pathways, lipid flux and storage, and mitochondrial function, with an accent on comparing and contrasting normal and pathologic metabolic states. Short talks were chosen from abstracts to increase the exposure of younger investigators and to highlight hot topics that complemented and extended the exciting program. There was ample time for discussion, which was very lively and often spilled over to other venues, including the cafeteria and social events designed to encourage interactions between trainees, young investigators, and senior faculty. In all, there were 40 talks—by speakers

Outdoor poster session

A. Marat, J. Walters

S. Hoffmann, A.K. Hopp

from Canada, Europe, and Asia in addition to the United States. Twelve of the talks were given by women.

In addition to the eight oral sessions, there were two lively poster sessions, featuring a total of 90 posters, that were extremely well attended and presented. All 10 sessions were characterized by open and wide-ranging discussions, and the meeting provided a unique forum for the exploration of the commonalities and differences in metabolic principles and details across different laboratories, systems, and diseases. All attendees gained in-depth exposure to the remarkable cell, organ, and disease specificity of metabolic flux and its regulation. Indeed, a great success was the interactive nature of the meeting, through which stimulating questions and discussion led to new concepts and future collaborations.

PROGRAM

Nucleus: Genetics/Genomes/Networks
Chairperson: M. Lazar, *University of Pennsylvania Perelman School of Medicine, Philadelphia*

Keynote Speaker
W.G. Kaelin, Jr., *HHMI/Dana-Farber Cancer Institute and Brigham and Women's Hospital, Boston, Massachusetts*

Signaling Control of Cellular Metabolism
Chairperson: M. Haigis, *Harvard Medical School, Boston, Massachusetts*

Signaling by Metabolites I
Chairperson: A. Pospisilik, *Max Planck Institute of Immunology and Epigenetics, Freiburg, Germany*

Mitochondria: Energetics
Chairperson: E. Taylor, *University of Iowa, Iowa City*

Mitochondria: Beyond Energetics
Chairperson: A. Walker, *University of Massachusetts Medical School, Worcester*

Integrated Physiology of Metabolism
Chairperson: J. Ferrer, *Imperial College London, United Kingdom/CIBERDEM, Spain*

B. Manning, W. Feng

Signaling by Metabolites II
Chairperson: W.L. Kraus, *University of Texas Southwestern Medical Center, Dallas*

Cancer Cell Metabolism
Chairperson: M. Loeken, *Joslin Diabetes Center, Boston, Massachusetts*

Retroviruses

May 22–27 349 Participants

ARRANGED BY **Vinay Pathak,** National Cancer Institute, Bethesda, Maryland
Stefan Sarafianos, University of Missouri School of Medicine, Kansas City

S. Sarafianos, V. Pathak

The organizers of this meeting—Drs. Vinay Pathak and Stefan Sarafianos—are regular attendees and view this meeting as the "intellectual home" in which they matured as scientists. This meeting has historically attracted many retrovirologists who attend the meeting regularly, which contributes to the sense that they belong to a community and to the formation of friendships and collaborations that have lasted for several decades.

The organizers selected Dr. Stephen Hughes, a senior retrovirologist at the National Cancer Institute, as one of the keynote speakers, and Dr. Daria Hazuda, Vice President, Infectious Diseases Discovery and Chief Scientific Officer, Merck Research Laboratories, as the second keynote speaker. Dr. Hughes, who has attended the meeting for decades, provided an overview of his many accomplishments and contributions to retrovirology over the years. Dr. Hazuda was the key person responsible for development of HIV-1 integrase inhibitors, and she provided an overview of events that led to the development of this new class of anti-AIDS drugs from the perspective of a senior scientist working at a pharmaceutical company. Both keynote addresses were well received and considered inspirational by the attendees. The organizers felt that Dr. Hazuda's keynote address was particularly important, as her success in industry hopefully served as an inspirational role model for female scientists.

Three annual prizes, originally instituted and endowed by the community of regular attendees with assistance from CSH Meetings staff, were also awarded. These prizes memorialize three of our former colleagues who passed away unexpectedly in recent years (the Daniel Wolf Prize, the Uta von Schwedler Prize, and the Andy Kaplan Prize) and serve to encourage up-and-coming scientists at the graduate and postdoc/junior levels. The prizes are also a testament to the community

A. Imle, W. Mothes

C. Carter, J. Skowronski

E. Rensen, G. Bedwell

C. Kieffer, T. Boecking

nature of this meeting and the continuity of the attendees over the years. The award recipients this year were Melissa Kane (Andy Kaplan Prize), Jordan Becker (Uta von Schwedler Prize), and Susana Valente (Daniel Wolf Prize).

The meeting was organized into 13 sessions (10 oral and three poster); the presentations featured unpublished findings, with an emphasis on basic research studies. The sessions utilized an integrated approach, focusing on traditional areas, such as specific aspects of the retrovirus life cycle, as well as on different types of host-cell defense mechanisms—in particular, specific restriction factors and the viral proteins that counteract them. Sessions on Virus Entry and IFITMs and a diverse session that included Evolution/Endogenous Retroelements, Vpr, Innate Immune Response, and Pathogenesis opened and closed the meeting, respectively. The organizers encouraged as many groups as possible to give talks by purposefully restricting each laboratory to no more than two talks at the meeting (and these two talks could not both be in the same session). In addition, the organizers introduced Short Talks, which consisted of a brief 5-minute presentation with 2 minutes for questions, to give more laboratories the opportunity to present their work in the oral sessions; the Short Talk investigators also presented a poster. This new feature was well received by the attendees. In addition, the organizers had three session chairs for four of the sessions, which gave some junior faculty members an opportunity to chair sessions and lead discussions. Chairing a session at this meeting is one indication of international recognition; the organizers wanted to give some junior faculty this opportunity to facilitate their transition to tenured faculty. Furthermore, there was a good balance of male and female co-chairs.

Scientific highlights of the meeting included the importance of conformational states of HIV-1 Env trimer structures, cryo-EM structures of supramolecular assemblies of lentiviral integrases, imaging of viral complexes in infected cells, insights into viral core uncoating both in vitro and in vivo, nuclear envelope docking, and nuclear import of viral complexes. A range of basic information regarding the retroviral life cycle was also presented. As previously, the meeting also exemplified new techniques and technologies that could be applied both to the discipline of retrovirology and to other areas of research.

Feedback from meeting attendees was uniformly positive; 85%–95% of the attendees rated the meeting as excellent to good, enjoyed the meeting, and were happy with the sessions and structure of the scientific program; 88% of the attendees were very satisfied or fairly satisfied with the Short Talks, confirming that this newly introduced format was a success.

The meeting was supported with funds from National Institute of Allergy & Infectious Diseases and ViiV Healthcare.

PROGRAM

Virus Entry and IFITMs
Chairpersons: W. Mothes, *Yale University, New Haven, Connecticut*; S. Neil, *King's College London, United Kingdom*

Restriction Factors I: SERINC5 and APOBEC3
Chairpersons: C. Goffinet, *TWINCORE, Hannover, Germany;* Y-H. Zheng, *Michigan State University, East Lansing*

Sixth Annual Uta von Schwedler Prize for Retrovirology
Awarded to J. Becker, *University of Wisconsin, Madison*

Reverse Transcription
Chairpersons: R. Ishima, *University of Pittsburgh School of Medicine, Pennsylvania;* B. Kim, *Emory University, Atlanta, Georgia*

Keynote Speaker
S. Hughes, *National Cancer Institute, Bethesda, Maryland*

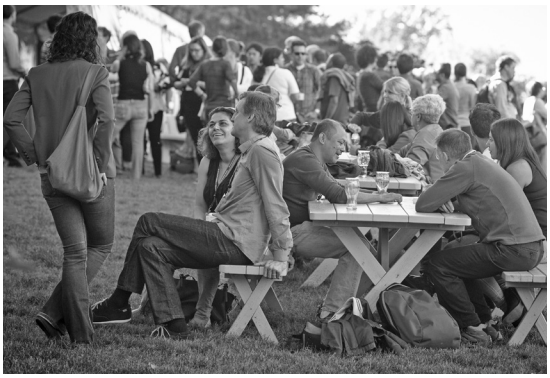
Uncoating and Nuclear Import
Chairpersons: G. Melikian, *Emory University, Atlanta, Georgia;* P. Zhang, *University of Pittsburgh School of Medicine, Pennsylvania*

Eleventh Annual Andy Kaplan Prize
Awarded to M. Kane, *The Rockefeller University, New York, New York*

Integration
Chairpersons: P. Cherepanov, *The Francis Crick Institute, London, United Kingdom;* D. Lyumkis, *The Salk Institute for Biological Studies, La Jolla, California*; K. Yoder, *The Ohio State University, Columbus*

RNA Expression, Transport, and Packaging
Chairpersons: F. Mustafa, *UAE University College of Medicine, Al Ain, United Arab Emirates;* A. Rice, *Baylor College of Medicine, Houston, Texas*

Assembly and Restriction Factors II: TRIM5a and TETHERIN/BST-2
Chairpersons: C. Carter, *Stony Brook University, New York;* O. Pornillos, *University of Virginia, Charlottesville*



Lunch on the lawn

Keynote Speaker
D. Hazuda, *Merck Research Laboratories, Kenilworth, New Jersey*

Restriction Factors III: SAMHD1 and New Restriction Factors/Mechanisms
Chairpersons: Y. Xiong, *Yale University, New Haven, Connecticut;* M. Roth, *Rutgers University-RWJMS, Piscataway, New Jersey;* K. Strebel, *NIAID, National Institutes of Health, Bethesda, Maryland*

Transmission and Latency
Chairpersons: O. Fackler, *University Hospital Heidelberg, Germany;* F. Maldarelli, *NCI-Frederick, Maryland;* A. Ono, *University of Michigan Medical School, Ann Arbor*

Seventh Annual Daniel Wolf Prize
Awarded Best Poster Presentation
to S. Valente, *Scripps Research Institute, Jupiter, Florida*

Evolution/Endogenous Retroelements, VPR, Innate Immune Response, and Pathogenesis
Chairpersons: J. Dudley, *University of Texas, Austin;* J. Skowronski, *Case Western Reserve School of Medicine, Cleveland, Ohio*

Genome Engineering: The CRISPR-Cas Revolution

July 21–23 374 Participants

ARRANGED BY **Jennifer Doudna,** HHMI/University of California, Berkeley
Maria Jasin, Memorial Sloan Kettering Cancer Center, New York, New York
Stanley Lei Qi, Stanford University, California
Jonathan Weissman, HHMI/University of California, San Francisco

J. Bondy-Denomy, R. Jaenisch

This genome engineering meeting was the third consecutive conference in the series. Genome engineering (gene editing) involving the introduction of breaks in the DNA backbone has been possible for several years. The discoveries of bacterial adaptive immunity and the co-opting of Cas9 and related components for programmable DNA recognition and cleavage have led to the widespread application of genome engineering, such that scientists working in many research areas and organisms can readily use the approach. Furthermore, Cas9 allows multiplexing and high-throughput genetic engineering to an extent that was not possible with previous approaches, and it is readily adapted for control of transcription, epigenetics, high-throughput mapping of genomics and cellular functions, rapid generation of genetically modified plants or animals, and gene drive for population control, among other applications, which were also included at the conference. The session topics are listed below in the Program.

We also invited Richard Hynes to give a talk on the ethical evaluation of applying CRISPR for clinical use of gene editing in humans, given the broad interest and progress in related fields. A goal of this meeting was to bring together researchers working in diverse fields to stimulate discussions and ideas to further exploit CRISPR-Cas9 and related technologies for biological discovery, organismal engineering, and medical applications. Twenty-four speakers were invited to cover these diverse topics. Another 27 speakers were chosen from submitted abstracts. Speakers represented institutions from the United States and abroad (United Kingdom, Canada, Italy, and China), with a few representatives from industry. Speakers were chosen from submitted abstracts

M. Jasin, J. Weissman, J. Doudna, S.L. Qi

A. Sfeir, A. Ciccia

B. Adamson, B. Conklin

M. Muzumdar, C-L. Wang

and included lab heads/staff scientists, postdoctoral fellows, and graduate students. Approximately 100 posters were presented in two sessions, complementing the oral presentations. The participants were from both academia and industry. Among approximately 374 participants, 105 (about one-third) were from industry.

The first session of the meeting focused on CRISPR biology with talks from Jennifer Doudna, Luciano Marraffini, and Eugene Koonin. Talks throughout the meeting utilized molecular, cell, and computational biology in diverse model organisms, as well as economically important and some unconventional organisms. Much of the data presented was unpublished or only very recently published.

This meeting was funded in part by Agilent Technologies; Advanced Analytical Technologies, Inc. Synthego; and Twist Bioscience.

PROGRAM

CRISPR
Chairpersons: E. Sontheimer, *University of Massachusetts Medical School, Worcester*; L. Marraffini, *The Rockefeller University, New York, New York*

Keynote Speaker
R.O. Hynes, *HHMI/Massachusetts Institute of Technology, Cambridge*

CRISPR Medicine
Chairpersons: C. Gersbach, *Duke University, Durham, North Carolina;* M. Porteous, *Stanford University, California*

Screens and Technology
Chairpersons: J. Shendure, *University of Washington, Seattle;* D. Liu, *Harvard University, Cambridge, Massachusetts*

Cancer and Stem Cells
Chairpersons: D. Huangfu, *Sloan Kettering Institute, New York, New York;* R. Jaenisch, *Whitehead Institute for Biomedical Research, Cambridge, Massachusetts*

Repairing DNA Breaks
Chairpersons: A. Sfeir, *New York University School of Medicine, New York;* J. Haber, *Brandeis University, Waltham, Massachusetts*

Gene Drive and Model Organisms
Chairpersons: A. Burt, *Imperial College London, United Kingdom;* B. Wiedenheft, *Montana State University, Bozeman*

Cell Death

August 15–19 168 Participants

ARRANGED BY **David Andrews,** University of Toronto, Canada
Anthony Letai, Dana-Farber Cancer Institute, Boston, Massachusetts
Karen Vousden, The Beatson Institute for Cancer Research, Glasgow, United Kingdom

D. Andrews, A. Letai, K. Vousden

This 12th Cell Death meeting was regarded by the attendees as outstanding. More than 130 excellent presentations and lively discussions centered on the topics listed below in the Program.

The keynote lectures were given by Craig Thompson on Manipulating Cell Death: The Discovery and Preclinical Development of BCL-2 Inhibitors, and Andrew Roberts on Clinical Progress with BCL2 Inhibition in Hematological Malignancies. These outstanding lectures were something of a victory lap for the last 20 years of this meeting, as many of the advances that led to successful clinical trials of BCL-2 inhibition in cancer came from this meeting. Essentially all key participants in this landmark achievement of translation to the clinic have attended this meeting. From two people who had lived it, we learned of the BCL-2 family from its inception as a breakpoint in a chromosomal translation in lymphoma (Craig Thompson) to its exploitation in the form of spectacularly successful clinical trials in cancer. We saw how history was made, not suddenly, but by the slow building of a scientific foundation that led to clinical advances whose potential is still largely unrealized.

It is hard to pick out individual talks when there were so many of high quality, but other highlights included presentations by Judy Campisi and Dahong Zhou on senolytic agents and the connection between aging and programmed cell death; Marc Tessier Lavigne on localized cell death in axons; Daniela Stooehr on how TRAIL-resistant cells can physically shield TRAIL-sensitive tumor cells; Darren Philips on a novel agent to target the TRAIL receptor; Ana García-Sáez with some beautiful physical biochemistry detailing BCL-2 family member recruitment to the mitochondrial membrane; and Scott Dixon on better representations of cell death data that incorporate

P. Juin, S. Fulda

X. Li, A.J. García-Sáez

kinetics, and the lethal fraction curve. Of the oral presentations, 38% were given by women, reflecting well the overall composition of the meeting.

J. Zhang, R. Crowder

There was definitely more of a translational bent to this meeting than ever before. This reflected the interests of the participants, but also the reality that cell death scientific achievements are being exploited in the clinic more than ever. By limiting the total number of talks, discussion remained lively and did not have to be curtailed, and sessions were completed more or less on time. Many commented that this made the meeting less stressful than previous meetings and more productive in terms of exploration of new ideas. An innovation suggested by Sandy Zinkel was incorporated on the fly—called lightning talks—where the presenters of 10 posters chosen for their excellence had exactly 1 minute and one slide to gain some additional attention. We think that all the participants and audience members found this to be a very informative and exciting 10 minutes! It will certainly be incorporated in future meetings as two separate lightning sessions, one preceding each poster session; 79 posters were also presented in two sessions, to crowded rooms.

The field of Cell Death is hitting its stride. It continues to re-invent itself at a basic level, with many different programmed cell death pathways now represented. We no longer need to look with expectation to the future for enormous societal benefits of this research, as its benefit to patients is now well recorded in clinical history. This meeting has proven in the past to be a worthy fulcrum on which exceptional basic science is demonstrably leveraged to clinical translation. We look forward eagerly to the great advances that will no doubt thrill us at the next meeting.

This meeting was funded in part by the National Institute on Aging, a branch of the National Institutes of Health; Agilent; and AstraZeneca.

PROGRAM

Senescence and Aging

Chairpersons: J. Campisi, *Buck Institute for Research on Aging, Novato, California;* D. Zhou, *University of Arkansas for Medical Sciences, Little Rock*

Neuronal Cell Death

Chairperson: M. Tessier-Lavigne, *Stanford University, California*

K. Newton, J. Zhang, A. Strasser, X. Lu

Cancer Cell Death
Chairperson: L. Walensky, *Dana-Farber Cancer Institute, Boston, Massachusetts*

Death Receptors, IAPs, and Necrosome
Chairpersons: K. Newton, *Genentech, South San Francisco, California;* D. Vucic, *Genentech, South San Francisco, California*

BCL-2 Family Proteins
Chairpersons: A.J. García-Sáez, *University of Tübingen, Germany;* J. Prehn, *Royal College of Surgeons in Ireland, Dublin*

Cell Death and Disease
Chairpersons: M. Blasco, *Spanish National Cancer Research Center, Madrid, Spain;* X. Lu, *University of Oxford, United Kingdom*

Novel Cell Death Regulation
Chairpersons: E. Cheng, *Memorial Sloan Kettering Cancer Center, New York, New York;* S. Martin, *Trinity College Dublin, Ireland*

Autophagy and Nonapoptotic Cell Death
Chairpersons: S. Dixon, *Stanford University, California;* E. White, *Rutgers Cancer Institute of New Jersey, New Brunswick*

K. Salnikow, E. Cheng

Model Systems
Chairperson: A. Strasser, *Walter & Eliza Hall Institute of Medical Research, Parkville, Australia*

Mitochondrial Cell Death
Chairpersons: X. Wang, *National Institute of Biological Sciences, Beijing, China;* F. Shao, *National Institute of Biological Sciences, Beijing, China*

Eukaryotic mRNA Processing

August 22–26 330 Participants

ARRANGED BY **Jean Beggs,** University of Edinburgh, United Kingdom
Alberto Kornblihtt, University of Buenos Aires, Argentina
Jens Lykke-Andersen, University of California, San Diego

A. Kornblihtt, J. Beggs, J. Lykke-Andersen

The 10th Eukaryotic mRNA Processing meeting was held this summer to present and discuss recent developments in mRNA metabolism in eukaryotes. This year, a highlight in the splicing structures and mechanisms session was the presentation of the structures of the budding yeast and human spliceosome B complexes obtained at atomic resolution by cryo-electron microscopy (also illustrated beautifully on the front cover of the abstract book). Comparison of these new structures with recently reported structures of other complexes (illustrated on the back cover of the abstract book) reveals the molecular rearrangements required for B complexes to transition to active spliceosomes. There were major advances in the characterization of 5′ splice site recognition in the yeast and human systems. Colocalization single-molecule spectroscopy (CoSMoS) was used to define the rules of 5′ splice site recognition by yeast U1 snRNP. This allowed understanding of why so many 5′ splice site sequences are recognized by the same factor and how sequence specificity and affinity of base pairing between U1 RNA and the 5′ splice sequence of the pre-mRNA are regulated by the presence of the protein components in the U1 snRNP particle. The meeting also offered space to reports that investigated the mechanisms of cytoplasmic splicing that participate in the unfolded protein response.

As in the past, the regulation of alternative pre-mRNA splicing was a key focus, helping to elucidate the importance of splicing in development and how splicing defects are related

H. Chen, F. Ibrahim

G. Singh

M. Chi, M. Hentze

L. Maquat, B. Hogg

to many genetic disorders and diseases. We heard about important developments from the ENCODE project that reveal the interactions and other characteristics of many RNA-binding proteins involved in different aspects of mRNA processing—a valuable resource for the whole community. There was an impressive demonstration of state-of-the-art single-molecule visualization techniques to track individual RNA molecules in real time in vivo, revealing their unexpected behavior during translation. Analyses of microRNAs, long noncoding RNAs, RNA modifications, virus-encoded RNAs, and the molecular basis of RNA-related diseases were also well represented.

In sessions on RNA turnover and translation, several exciting presentations focused on the connection between these two processes. Highlights included talks on the effects of codon-dependent translation efficiency on mRNA stability and poly(A)-tail length, including competing models of underlying mechanisms. There were also new insights presented into mechanisms of mRNA quality control as well as talks on global, structural, and microscopy approaches to monitor translation and mRNA decay. In the session on noncoding RNAs, a highlight was evidence presented that long-known enigmatic vault RNAs serve as riboregulators in autophagy. In addition, exciting presentations focused on roles of small RNAs in mRNA regulation and on pathways serving to prevent accumulation of long noncoding RNAs in the cytoplasm, where they could interfere with the translation machinery.

There was an exciting session on functional coupling of cotranscriptional splicing with other nuclear processes. Long-read sequencing of nascent transcripts in fission yeast revealed that most were either fully spliced or unspliced, with few being partially spliced, suggesting a potential cooperativity in splicing neighboring introns. The link between splicing and histone H3K36 trimethylation was investigated in budding yeast, using an auxin degron to knock down splicing factors at sequential steps in the splicing pathway. Recruitment of the Set2 H3K36 methyltransferase seems to require cotranscriptional complex A formation, but its full methyltransferase activity requires further spliceosome assembly. It was shown that slowing down transcription affects the secondary structure of the transcript, which, in the case of histone genes, leads to the generation of abnormal polyadenylated histone mRNAs.

In the sessions devoted to RNA processing and disease, interesting discussions were raised on how mutations in splicing factors cause myelodysplastic syndromes. In particular, evidence was presented to consider that besides alteration of alternative splicing, splicing factor mutation leads to R-loop formation that in turn causes genotoxic stress. Another topic of interest was the study of how disruption of neuronal alternative splicing networks is key for autism spectrum disorders. It will be clear from this brief overview that recognition of the influence of RNA in multiple cellular processes continues to grow!

T. Johnson, M. Ares, K. Neugebauer

PROGRAM

RNA Turnover and Quality Control
Chairpersons: A. Corbett, *Emory University School of Medicine, Atlanta, Georgia;* L. Maquat, *University of Rochester Medical Center, New York*

Splicing Structures and Mechanisms
Chairpersons: B. Schwer, *Weill Cornell Medical College, New York, New York;* J. Staley, *University of Chicago, Illinois*

RNA–Protein Interactions
Chairperson: G. Yeo, *University of California, San Diego*

3′-End Processing and Poly(A)
Chairperson: S. West, *University of Exeter, United Kingdom*

Splicing Regulation
Chairpersons: C. Smith, *University of Cambridge, United Kingdom;* J. Valcarcel, *Center for Genomic Regulation (CRG), Barcelona, Spain*

Noncoding RNAs
Chairpersons: B. Bass, *University of Utah, Salt Lake City;* N. Kim, *Seoul National University, Korea*

Cotranscriptional RNA Processing and RNA Modification
Chairpersons: D. Bentley, *University of Colorado School of Medicine, Aurora;* K. Neugebauer, *Yale University, New Haven, Connecticut*

Translation
Chairpersons: G. Chanfreau, *University of California, Los Angeles;* M. Hentze, *European Molecular Biology Laboratory, Heidelberg, Germany*

RNA Processing in Development and Disease
Chairpersons: J. Steitz, *HHMI/Yale University, New Haven, Connecticut;* X-D. Fu, *University of California, San Diego*

Mechanisms of Eukaryotic Transmission

August 29–September 2 378 participants

ARRANGED BY **Patrick Cramer,** Max Planck Institute, Münich, Germany
John Lis, Cornell University, Ithaca, New York
Jane Mellor, University of Oxford, United Kingdom

J. Lis, J. Mellor, P. Cramer

Regulation of gene transcription plays a central role in the growth and development of eukaryotic organisms. Transcriptional responses occur as a consequence of cell signaling, environmental stresses, and developmental cues. The field of transcription encompasses a broad range of study from structural biology to developmental biology and functional genomics. This 15th meeting brought together a diverse group of scientists and appropriately covered many aspects of the field in eight plenary sessions and two poster sessions.

The meeting was structured based on steps in the cycle of transcription, and it began with a session that described the initial step of modulating access of transcription factors and RNA polymerase II (Pol II) to chromatin-packaged DNA. The second session included exciting mechanistic studies of transcription initiation and cryo-EM structure studies of human and yeast Mediator and Pol II initiation complexes. Another session (fourth session) revealed new insights into the regulation occurring at promoter-proximal pause release. The fifth session covered the regulatory interplay of noncoding and coding transcription and chromatin modification, and the one following described new insights into long-range interactions of regulatory elements and promoters. Another plenary session (eighth session) covered cotranscriptional RNA processing and termination, and how transcription studies provide for mechanistic understanding and treatment of many diseases was covered on the ninth session. The 10th and final session held a large fraction of participants through to the last talk of the meeting. The session topics are listed below in the Program.

Interspersed with these oral presentations were the two evening poster sessions (Sessions 3 and 7), where a wide variety of exciting unpublished transcriptional research was presented. The posters

N. Proudfoot, S. Churchman

A. Angel, T. Brown

B. Martin, B. Plosky

M. Eilers, E. Wolf

were extremely thoughtful and covered state-of-the-art research, and the evening format provided a relaxed atmosphere for discussions. A "career event" dinner allowed established investigators to share their insights in informal discussions with younger participants.

This meeting was extremely well received by the participants and will occur again in 2019.

PROGRAM

Creating Chromatin Architecture and Modulating Accessibility of Promoters and Enhancers
Chairperson: K. Adelman, *Harvard Medical School, Boston, Massachusetts*

Mechanistic Roles of Transcription Factors I: Early Steps
Chairperson: J. Ahringer, *Wellcome Trust/Cancer Research UK Gurdon Institute, Cambridge, United Kingdom*

Mechanistic Roles of Transcription Factors II: Pause Release and Elongation
Chairperson: S. Buratowski, *Harvard Medical School, Boston, Massachusetts*

The Regulatory Interplay of Noncoding and Coding Transcription and Chromatin Modification
Chairperson: R. Dowell, *University of Colorado, Boulder*

Enhancers, Genome Architecture, and Long-Range Regulatory Interactions in Development
Chairperson: E. Furlong, *EMBL, Heidelberg, Germany*

Cotranscriptional RNA Processing and Termination
Chairperson: M. Levine, *Princeton University, New Jersey*

Transcription Regulation and Medicine
Chairperson: D. Libri, *Institut Jacques Monod, Paris, France*

New Technologies and Concepts in Studies of Transcription Regulation
Chairperson: B.F. Pugh, *Pennsylvania State University, University Park*

J. Reese, A. Shilatifard

Eukaryotic DNA Replication and Genome Maintenance

September 5–9 344 Participants

ARRANGED BY **Karlene Cimprich,** Stanford University, California
Anne Donaldson, University of Aberdeen, United Kingdom
Anindya Dutta, University of Virginia, Charlottesville

A. Donaldson, A. Dutta, K. Cimprich

This 16th biennial meeting provided a pivotal opportunity for discussion and exchange of ideas in the rapidly advancing field of chromosome duplication and genome stability research. Delegates attended from 22 different countries, reflecting the diversity, quality, and breadth of this research area, with investigators keen to share and discuss the latest breakthroughs. Approximately half of the 344 attendees were Ph.D. students or postdoctoral scientists. The number and diversity of attendees are underpinned by this meeting's established position as the foremost conference in the field of DNA replication. There were 80 platform talks grouped in eight sessions, with a total of 261 platform and poster presentations. Every one of the talks drew stimulating questions from the audience and provoked ongoing discussion continuing into the social parts of the program. The poster sessions were packed throughout and distinguished by lively discussion, with interactions encouraged by the welcoming design of the Nicholls-Biondi poster pavilion. A new feature that was well received by the attendees was that both poster sessions were held in the evenings.

The 10 platform sessions started with broader questions on genome instability and human disease and ended with molecular details such as the response of the cell to fork stalling and DNA damage; the sessions are listed below in the Program. Early sessions described links between genome instability, mutagenesis, disorders of replication, and human disease. The intermediate session focused on how the chromatin affects replication and how the initiator proteins initiate origin firing. The later sessions discussed both how the cell responds to fork stalling and replication stress and how the replication apparatus responds to transcription and DNA

T. Miller, G. Coster

A. Vindigni, J. Mendez

J. Peake, E. Noguchi

W. Feng, T. Pohl

damage. Interesting advances included the demonstration that genome instability can produce DNA that stimulates the innate immune response, implication of unexpected proteins such as the proteasome shuttle proteins or Drosha in the DNA-damage response following replication stress or DNA breaks, the presence of genetic polymorphisms that affect replication timing of adjoining parts of the genome, and the use of an in vitro system of replication initiation to demonstrate that MCM10 directly activates the helicase activity of the CMG complex. R-loops showed up in different contexts ranging from genomic instability caused by the collision of replication fork with an R-loop to the role of R-loops in promoting chromosome segregation in mitosis. A few controversies also appeared, such as the suggestion that some mammalian cancer cells can survive without a subunit of ORC and whether G4 quadruplexes are functional elements of origins in mammalian cells.

The strategy of broadening the conference scope to encompass the intersection of replication processes with genome stability mechanisms is working well and is expanding the scope of research covered. New technologies to track replication by advanced microscopy or using single-molecule studies are also being adopted in more laboratories.

Overall, the meeting provided an excellent update on current understanding, with the broad range and high quality of scientific advance and discussion fully confirming this as the preeminent conference in this fast-moving and dynamic research field.

This meeting was funded by the National Cancer Institute, a branch of the National Institutes of Health, and Genomic Vision.

PROGRAM

Genome Instability and Mutagenesis
Chairpersons: J. Rouse, *University of Dundee, United Kingdom;* A. Smogorzewska, *The Rockefeller University, New York, New York*

DNA Replication and Human Disease
Chairpersons: P. Pasero, *Institute of Human Genetics-CNRS, Montpellier, France;* A. Jackson, *MRC Human Genetics Unit, Edinburgh, United Kingdom*

Chromatin and Replication Timing
Chairpersons: R. Duronio, *University of North Carolina, Chapel Hill;* M. Gómez, *Centro de Biología Molecular Severo Ochoa, Madrid, Spain*

T. Tsurimoto, S. Bell

Replication Initiation at Origins
Chairpersons: J. Diffley, *Francis Crick Institute, United Kingdom;* H. Li, *Van Andel Research Institute, Grand Rapids, Michigan*

Coordinating the Cell Cycle and DNA Replication; Other Initiation Systems
Chairpersons: D. Branzei, *IFOM-FIRC Institute of Molecular Oncology, Milan, Italy;* J. Li, *University of California, San Francisco*

Replisomes and Fork Progression
Chairpersons: B. Tye, *Hong Kong University of Science and Technology, Hong Kong;* T. Tsurimoto, *Kyushu University, Fukuoka, Japan*

Fork Stalling and Replication Stress
Chairpersons: E. Brown, *University of Pennsylvania, Perelman School of Medicine, Philadelphia;* J. Walter, *HHMI/Harvard Medical School, Boston, Massachusetts*

Response of Replication to Transcription and DNA Damage
Chairpersons: M. Lopes, *University of Zürich, Switzerland;* J. Sale, *MRC Laboratory of Molecular Biology, Cambridge, United Kingdom*

Microbial Pathogenesis and Host Response

September 12–16 296 Participants

ARRANGED BY **Denise Monack,** Stanford University, California
Raphael Valdivia, Duke University Medical Center, Durham, North Carolina
Malcolm Whiteway, Concordia University, Canada

M. Whiteway, D. Monack, R. Valdivia

Despite advances in modern healthcare, infectious diseases continue to be major causes of human morbidity and mortality. The evolution of microbial pathogens with humans has resulted in complex interactions that impact the struggle between the infectious invader and the susceptible host. Increased understanding of these interactions with the goal of developing new therapeutics and preventive strategies requires collaborative and interdisciplinary scientific approaches. This meeting brought together a diverse group of international scientists who approach the study of bacterial, parasitic, and fungal pathogens from a broad range of perspectives. Investigators from the disciplines of molecular microbiology, eukaryotic cell biology, immunology, and genomics and representing academia, scientific publishing, industry, and the public health sector shared recent findings concerning microbial and host aspects of infectious diseases.

The meeting focused on the cross talk between microbial pathogens and the host, facilitating an increased understanding of host response and defense mechanisms to these invading microbes. Oral sessions were topic-based and included studies of multiple and diverse organisms (see Program below for session topics). Speakers for each session were a mixture of established leaders in the field and young investigators. Half of the speakers, including postdoctoral fellows and graduate students, were chosen from submitted abstracts. Dr. Reginne Kahmann, an internationally recognized leader in the field of plant fungal pathogenesis, presented the keynote address. Her presentation, Understanding Fungal Plant Pathogens—From Secreted Effectors to New Targets for Disease Intervention, provided an exciting story of the complexity of plant/fungal pathogen interactions and drew fascinating parallels between fungal and bacterial effector systems.

E. Rosowski, J. Saeij

D. Lenehan, A. Schulz

E. Hinson, M. Wallace

M. O'Riordan, L. Knodler, J. Elfenbein

The informal atmosphere combined with the broad perspectives of the meeting participants resulted in a free flow of novel and refreshing ideas on pathogenesis and clinical treatment, with the atmosphere of a small meeting. Active questioning and discussion followed all oral presentations, was evident throughout the posters sessions, and continued during a wine and cheese reception and other social gatherings. We strongly encouraged submission of abstracts by junior researchers in the field, and many young investigators were in attendance. Some of these interactions have already produced fruitful scientific collaborations.

This meeting was partially supported by funds from the National Institute of Dental and Craniofacial Research and the National Institute of Allergy and Infectious Diseases.

PROGRAM

Visualizing Host–Pathogen Interactions
Chairperson: A. Sil, *University of California, San Francisco*

Hosts Responses and Defense
Chairperson: M. O'Riordan, *University of Michigan, Ann Arbor*

Nutrition Matters: Metabolism and Pathogenesis
Chairperson: J. Kronstad, *University of British Columbia, Vancouver, Canada*

Pathogen-Innate Immune System Interface
Chairperson: K. Cadwell, *New York University School of Medicine, New York*

Surfaces and Pathogenesis
Chairperson: C. Munro, *University of Aberdeen, United Kingdom*

Microbe–Microbe Interactions within Hosts
Chairperson: M. Blokesh, *Swiss Federal Institute of Technology Lausanne (EPFL), Switzerland*

Keynote Speaker
R. Kahmann, *Max Planck Institute for Terrestrial Microbiology, Marburg, Germany*

Stem Cell Biology

September 25–29 192 Participants

ARRANGED BY **Fiona Watt,** King's College London, United Kingdom
Marius Wernig, Stanford University, California
Kenneth Zaret, University of Pennsylvania, Philadelphia

K. Zaret, F. Watt, M. Wernig

Of the 192 participants at this fourth biannual conference, ~23% were graduate students and 25% were postdocs from 20 different countries. Representatives from major publishing houses were present such as *Science, Cell Press, Journal of Cell Biology, Nature, Cell Biology,* and *Development.* The organizers sought to emphasize cross-disciplinary work on cellular and molecular mechanisms of the stem cell and stem cell–related tissue systems. The meeting was thus organized in thematic topics, which are listed below in the Program. Speakers comprised a mix of internationally known leaders in the individual disciplines as well as emerging junior researchers, who presented their recent work. In addition, 18 short talks were chosen from abstracts featuring postdocs and students as well as some additional early career independent investigators. Main talks were 20–25 minutes, plus 5–10 minutes of discussion, whereas short talks were 10 minutes plus 5 minutes of discussion. The keynote lecture was given by Irv Weissman, who presented exciting unpublished work on hematopoietic stem cells in physiologic conditions and leukemia. Most of the presented work was unpublished and led to stimulating discussions immediately after the talk, during poster sessions, meals, and at the bar. Preliminary (informal) feedback by students, postdocs, speakers, and journal editors was uniformly very positive. Attendees liked the mixed topics of sessions and the fact that most speakers stayed for several days and were approachable throughout the conference.

This meeting was funded in part by Agilent Technologies.

I. Weissman, S. Lowell

A. Webb, B. Hogan

E. Passegue, P. Beachy

S. Reusch, L.J. Pilaz

PROGRAM

Keynote Speaker
I. Weissman, *Stanford University, California*

Establishing and Exiting the Stem Cell State
Chairperson: K. Hochedlinger, *Harvard Medical School, Massachusetts General Hospital, Boston*

Induced Reprogramming
Chairperson: S. Lowell, *MRC Centre for Regenerative Medicine, Edinburgh, United Kingdom*

New Advances in Stem Cell and Lineage Dynamics
Chairperson: J. Rajogopal, *Harvard Medical School, Massachusetts General Hospital, Boston*

Reconstituting Tissues In Vitro and in Disease
Chairperson: E. Passegue, *Columbia University Medical Center, New York*

Epigenetic Control of Stem Cell States
Chairperson: C. Lengner, *University of Pennsylvania, Philadelphia*

Mechanisms of Homeostatic Self Renewal
Chairperson: K. Plath, *University of California, Los Angeles*

Tissue Repair through Physiologic Reprogramming
Chairperson: V. Greco, *Yale University School of Medicine, New Haven, Connecticut*

Neurobiology of *Drosophila*

October 3–7 448 Participants

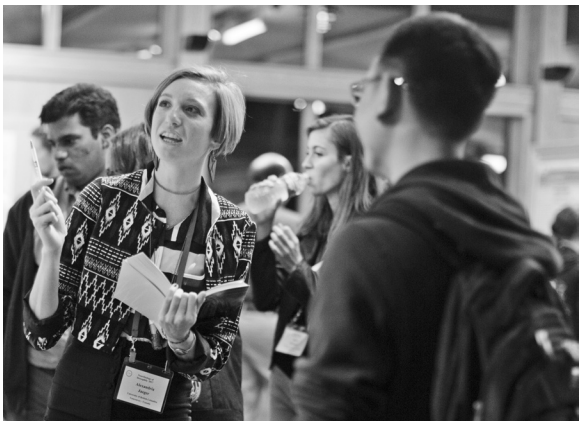
ARRANGED BY **Heather Broihier,** Case Western Reserve University, Cleveland, Ohio
J. Troy Littleton, Massachusetts Institute of Technology, Cambridge

N. Rangarajan

This meeting provided a forum for the discussion of new discoveries, techniques, and advances in *Drosophila* neurobiology. Eleven sessions ran in series over 4 days, with alternating platform and poster presentations. A special technology session with emphasis on new methods and approaches broadly applicable to the field was a new and well received addition. Alternating between the more and less formal presentations provided excellent opportunities to meet other attendees and discuss scientific issues one-on-one. These interactions were excellent for facilitating collaborations, exchange of reagents (e.g., antibodies, clones, mutants, and other stocks), methods (genetic, physiological, optical), and ideas between both new and established investigators. The eight platform session topics were chosen to reflect the areas in which cutting-edge advances are being made and are listed below in the Program. Session chairs and the meeting organizers selected presenters for these platform sessions from the submitted abstracts, and the remaining abstracts were presented as posters. The vast majority of the speakers were graduate students and postdoctoral fellows, and approximately half were female. Many people commented positively on the addition of the joint technology session, which allowed broad discussion of new techniques. In the keynote address, the Benzer Lecture, Hugo Bellen gave a stimulating presentation on the role of glial lipid droplets in neurodegenerative disease. The Elkins Award plenary lecture is presented at each meeting by a graduate student whose dissertation exemplifies the finest work in our field. This year the Elkins Lecture was presented by Dr. Raphael Cohn, who trained as a graduate student with Dr. Vanessa Ruta at Rockefeller University, and who presented his spectacular work on the role of neuromodulation in learning and memory in the *Drosophila* mushroom bodies. New this year was a reunion for all former participants of the *Drosophila* CSH meetings. This provided a forum that stimulated lively discussion and interaction among the many participants of prior courses. The meeting also had a discussion session with key lab members

Poster session in Nichols Biondi Pavilion

J. Vaughen, S. Lewis



A. Jaeger

who contributed to the 2017 Nobel Prize, which was awarded for circadian research using the *Drosophila* model. This bonus session allowed members to share in the key discoveries and the excitement that accompanied the awarding of the prize for *Drosophila* neuroscience research. This meeting was well attended, with presentations spanning the breadth of modern neurobiology. The many opportunities for interaction and career development fostered by this meeting are sure to enhance this vibrant field.

This meeting was funded in part by the National Institute of Neurological Disorders and Stroke, a branch of the National Institutes of Health.

PROGRAM

Higher Brain Function and Behavior
Chairperson: S. Waddell, *University of Oxford, United Kingdom*

Neuronal and Brain Development
Chairperson: E. Heckscher, *University of Chicago, Illinois*

Sensory Systems
Chairperson: V. Ruta, *The Rockefeller University, New York, New York*

Elkins Memorial Lecture Award Recipient
Raphael Cohn, *The Rockefeller University, New York, New York:* Circuit mechanisms for flexible sensorimotor processing in *Drosophila*.

Technological Innovations
Chairperson: P. Tomancak, *Max Planck Institute, Dresden, Germany*

J. Flyer-Adams, F. Rouyer

J. Meng, P. Iturralde

Neuronal Circuits
Chairperson: A Cardona, *HHMI/Janelia Farm Research Campus, Ashburn, Virginia*

Seymour Benzer Lecture
H.J. Bellen, *Baylor College of Medicine, Houston, Texas:* Mitochondrial dysfunction in neurons: Lipid synthesis in neurons versus lipid droplets in glia

Disease Models and Mechanisms
Chairperson: P. Verstreken, *VIB, KU Leuven, Belgium*

Synaptic Transmission and Plasticity
Chairperson: K. O'Connor-Giles, *University of Wisconsin, Madison*

Neuronal Cell Biology
Chairperson: A. Rodal, *Brandeis University, Waltham, Massachusetts*

T. Hafer, S. Sprecher

Biology of Cancer: Microenvironment and Metastasis

October 10–14 237 Participants

ARRANGED BY **Scott Lowe,** Memorial Sloan Kettering Cancer Center, New York, New York
Senthil Muthuswamy, Beth Israel Deaconess Medical Center, Boston
M. Celeste Simon, University of Pennsylvania Perelman School of Medicine, Philadelphia

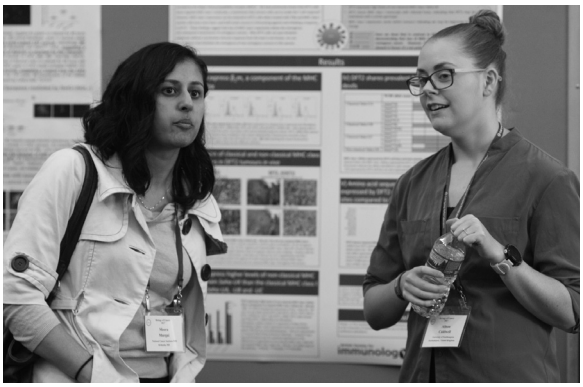
S. Muthuswamy, S. Lowe, M.C. Simon

Our understanding of cancer is evolving rapidly. We have now moved away from the early thoughts of cancer as a clonal growth of one rogue cell, to the recognition that cancer is a highly heterogeneous entity both at the level of genetic mutations and at the level of cell types within a growing tumor. Furthermore, we are now beginning to appreciate the fact that cancers not only metastasize to different parts of the body, but that they also have systemic effects that produce and influence pathologies in the organism as a whole. This meeting aimed to capture this complexity by considering the cancer problem from an integrated and organismal perspective. It highlighted emerging topics that are actively investigated by cancer researchers worldwide and brought together molecular, cellular, and computational biologists to discuss recent advances on a diverse array of topics such as tumor immunology, cellular plasticity, stress adaptation, metastasis, cancer dormancy, noncoding RNA, cell metabolism, and systemic effects. All of the talks were of extremely high caliber, which led to extended and insightful discussion from the audience. Joan Massagué, the keynote speaker, brilliantly spoke about the complexities of the metastatic process and discussed new results that highlighted the cooperation between cancer cell and noncancer cells during metastasis and dormancy. Additionally, in a new aspect of the meeting format, we incorporated a session aimed at highlighting and recognizing the outstanding science of early career investigators. Specifically, we invited eight investigators who are within seven years of starting their independent laboratories to present their most exciting work. The goal of this new initiative was not only to recognize these emerging stars, but also to serve as a motivation for trainees and other junior investigators at the meeting. Finally, to motivate trainees to stay engaged in cancer research, we incorporated a career mentoring session

J. Wells, A. Gamero

R. White, M. Yu

R. Piskol, B. Zetter



M. Murgai, A. Caldwell

that included an open discussion on the historical perspective of funding for cancer research. Both the mentoring sessions and the session highlighting emerging stars were extremely well received. Collectively, participants enjoyed both the diversity and the depth of the topics covered.

This meeting was funded in part by Agilent, with major support provided by Northwell Health.

PROGRAM

Junior Stars Session I
Chairperson: C. Abate-Chen, *Columbia University Medical Center, New York, New York*

Tumor Immunology
Chairpersons: M. Egeblad, *Cold Spring Harbor Laboratory*; T. Gajewski, *University of Chicago, Illinois*

Epigenome/Plasticity/Heterogeneity
Chairpersons: J. Gray, *Oregon Health & Science University, Portland;* B. Stanger, *University of Pennsylvania Perelman School of Medicine, Philadelphia*

Translational Control and Stress Adaptation
Chairpersons: D. Bar-Sagi, *New York University Langone Medical Center, New York;* W. Kaelin, *HHMI/Dana-Farber Cancer Institute, Boston, Massachusetts*

Careers in Science
Chairperson: B. Zetter, *Boston Children's Hospital, Massachusetts*

Metastasis/Detection/Dormancy
Chairpersons: L. Chodosh, *University of Pennsylvania Perelman School of Medicine, Philadelphia;* Y. Kang, *Princeton University, New Jersey*

Tumor Context and RNA
Chairpersons: D. Spector, *Cold Spring Harbor Laboratory*; P.P. Pandolfi, *Beth Israel Deaconess Medical Center/Harvard Medical School, Boston, Massachusetts*

Systemic Effects of Cancer
Chairpersons: D. Lyden, *Weill Medical College of Cornell University, New York;* D. Fearon, *Cold Spring Harbor Laboratory*

Joint Keynote Session
J. Massagué, *Memorial Sloan Kettering Cancer Center, New York, New York;* E. Mardis, *Institute for Genomic Medicine at Nationwide Children's Hospital, Columbus, Ohio;* T. Lynch, *Bristol-Myers Squibb, New York, New York*

W. Kaelin, J. Gray

Next-Generation Cancer Clinical Trials

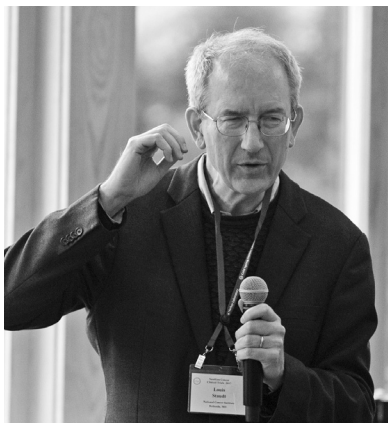
October 13–16 70 Participants

ARRANGED BY **Elizabeth Jaffee,** Johns Hopkins University, Baltimore, Maryland
Robert Maki, Northwell Health, New Hyde Park, New York
David Tuveson, Cold Spring Harbor Laboratory

P. Workman, D. Tuveson

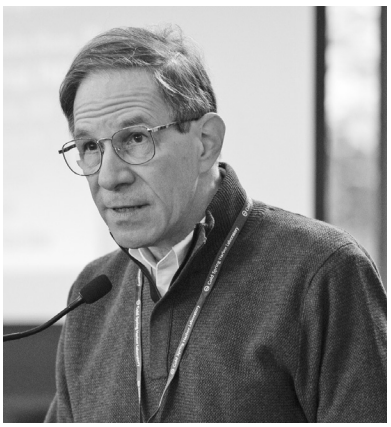
This inaugural meeting, held at the Nichols-Biondi Hall, addressed the increased application of basic cancer research to clinical problems at CSHL, highlighted by the recent affiliation of CSHL with Northwell Health, a regional healthcare provider caring for more than 19,000 new cancer patients a year. Seventy people attended this meeting, representing U.S., Canadian, and European leading authorities in cancer research. The goals of the meeting were to highlight recent clinical therapeutic and diagnostic advances and to discuss the emerging scientific and logistical obstacles that hinder dramatic improvements in patient outcome. The opening keynote addresses were given by Elaine Mardis, who discussed progress in cancer genetics that is shaping personalized medicine approaches, and Tom Lynch, who discussed the important role of industry working in partnership with academic groups for successful drug development. Twenty-three talks were given by leading physicians and scientists—eight given by women—focusing on the themes of diagnostics, therapeutics, and tumor immunology. In addition, there were two forums discussing The Role of Big Data, directed by Lou Staudt, and The Ecosystem of Cancer Therapeutic Discovery, directed by Paul Workman. A final keynote presentation was given by Irving Weisman on new methods to promote antitumor immunity. The discussion sessions were lively and resulted in a dynamic engagement of the audience. There were several collaborative projects proposed between faculty who were interacting for the first time and much enthusiasm to repeat this conference in the near future.

This meeting was funded by Northwell Health.



L. Staudt

S. Sanduja



S. Markowitz

R. Jain

I. Weissman

PROGRAM

Diagnostics and Therapies

M. Pomper, *Johns Hopkins University School of Medicine, Baltimore, Maryland*
S. Nimmagadda, *Johns Hopkins University, Baltimore, Maryland*
R. Jain, *Massachusetts General Hospital, Boston*
D. Paul, *Monter Cancer Center, New Hyde Park, New York*
S. Markowitz, *Case Western Reserve University, Cleveland, Ohio*

Forum I: Big Data

Chairperson: L. Staudt, *National Cancer Institute, Bethesda, Maryland*
Panelists: E. Mardis, *Washington University in St. Louis, Missouri*; S. Markowitz, *Case Western Reserve University, Cleveland, Ohio;* L. Siu, *Princess Margaret Hospital, Toronto, Ontario, Canada*

Therapeutics

A. Bardelli, *University of Torino School of Medicine, Italy*
V. Coppola, *The Ohio State University, Columbus*
J. Grandis, *University of California, San Francisco*
L. Staudt, *National Cancer Institute, Bethesda, Maryland*
P. Dirks, *Hospital for Sick Children, University of Toronto, Ontario, Canada*
D. Stover, *Ohio State University, Columbus/Dana-Farber Cancer Institute, Boston Massachusetts*
H. Wang, *St. Jude Children's Research Hospital, Memphis, Tennessee*
H. Tiriac, *Cold Spring Harbor Laboratory*
K. Stegmaier, *Dana-Farber Cancer Institute, Boston, Massachusetts*
L. Siu, *Princess Margaret Hospital, Toronto, Ontario, Canada*
A. Biondo, *The Institute of Cancer Research, London, United Kingdom*
K. Rai, *Feinstein Institute for Medical Research, Manhasset, New York*
P. Workman, *The Institute of Cancer Research, London, United Kingdom*

Forum 2: The Ecosystem of Cancer Drug Discovery

Chairperson: P. Workman, *The Institute of Cancer Research, London, United Kingdom*
Panelists: L. Staudt, *National Cancer Center, Bethesda, Maryland;* L. Siu, *Princess Margaret Hospital, Toronto, Ontario, Canada;* A. Bardelli, *University of Torino School of Medicine, Italy;* R. Maki, *Northwell Health, New Hyde Park, New York*

Session IV: Immunology

D. Le, *Sidney Kimmel Comprehensive Cancer Center, Baltimore, Maryland*
Y. Kanjanapan, *Princess Margaret Cancer Centre, Toronto, Ontario, Canada*
E. Jaffee, *Johns Hopkins University, Baltimore, Maryland*
M. Yarchoan, *Johns Hopkins University, Baltimore, Maryland*
S. Topalian, *Johns Hopkins University, Baltimore, Maryland*

Keynote Speaker

I. Weissman, *Stanford University, California*

40 Years of mRNA Splicing: From Discovery to Therapeutics

October 22–25 272 Participants

ARRANGED BY **M. Pollock,** Cold Spring Harbor Laboratory
P. Sharp, Massachusetts Institute of Technology, Cambridge
J. Steitz, HHMI/Yale University, New Haven, Connecticut

The Cold Spring Harbor Laboratory Archives and Genentech Center for the History of Molecular Biology and Biotechnology, together with the Meeting & Courses department, hosted this seventh meeting, which brought together some of the most important mRNA splicing researchers from the past and present.

Meeting speakers included investigators who made many of the seminal discoveries that began the field, as well as those who are working in the field now. Having the meeting at Cold Spring Harbor Laboratory was especially significant as critical mRNA splicing discoveries here earned Richard Roberts a share in the 1993 Nobel Prize in Physiology or Medicine, along with Phillip Sharp, who had been a Senior Scientist at the Laboratory in the early 1970s. Both Nobelists spoke at the meeting. A history session at the meeting included Cold Spring Harbor scientists Louise Chow and Richard Gelinas, who played critical roles in the splicing discovery. Also attending was newly minted Nobelist Michael Rosbash, who shared the 2017 Nobel Prize in Physiology or Medicine for "discoveries of molecular mechanisms controlling the circadian rhythm." He talked about his major contribution to the mRNA splicing field.

The discovery of mRNA splicing in 1977 established a new step in the central dogma of molecular biology—the "flow" of genetic information from DNA to RNA to the protein product. mRNA splicing is essentially the editing of the precursor mRNA transcript copied from nuclear DNA into a mature mRNA molecule that can then be translated into a protein. Key to this process is the spliceosome, an RNA–protein complex that catalyzes the splicing reactions. mRNA splicing has been revealed to be exquisitely complex as new technologies have evolved to study it in ever closer detail.

mRNA splicing plays a large role in a wide range of diseases, including motor neuron diseases, muscle degenerative diseases, retinopathies, and cancers, to name a few. Disruptions in the splicing

The Splicing group

P. Sharp, M. Konarska

M. Rosbash, C. Burge

process via mutation have been estimated to constitute as much as 50% of human disease mutations. Deciphering the mechanisms of these mutations is essential to developing effective therapies that can correct or inhibit the resulting pathological splicing events. Remedies being explored include antisense oligonucleotides, RNA-binding molecules, and chemicals that interact with splice site selection.

The main topics of discussion and the Chairpersons are listed below in the Program. Discussions encompassed not only what has been accomplished, but also how this history continues to shape the future of mRNA splicing research and its applications. Some of the most interesting and exciting current research in the field presented included therapeutics-based approaches to mRNA splicing in spinal muscular atrophy by Adrian Krainer and cryo–EM of the spliceosome by Kiyoshi Nagai.

Financial support for the meeting was provided by Major Sponsors Biogen and Moderna Therapeutics and Gold Sponsors Alnylam Pharmaceuticals and New England BioLabs. Additional support was provided by Ionis Pharmaceuticals.

PROGRAM

History and Overview
J. Darnell, *The Rockefeller University, New York, New York*
I. Mattaj, *EMBL, Heidelberg, Germany*

Biology of Spliceosome
Chairpersons: T. Nilsen, *Case Western Reserve University, Cleveland, Ohio;* M. Ares, *University of California, Santa Cruz*

Introns, Exons, and Alternative Splicing
Chairpersons: W. Gilbert, *Harvard University, Cambridge, Massachusetts;* C. Moore, *Tufts University School of Medicine, Boston, Massachusetts*

Spliceosome Complexes
Chairpersons: M. Rosbash, *Brandeis University, Waltham, Massachusetts;* W-Y. Tarn, *Academia Sinica, Taipei, Taiwan*

R. Darnell, B. Stillman, R. Roberts, A. Krainer

L. Chow, T. Broker

M. Casadio, X-D. Fu

M. Pollock, W. Gilbert, R. Vaz-Drago

Speakers
B. Stillman and J. Watson, *Cold Spring Harbor Laboratory*

Diseases of mRNA Splicing
Chairperson: S. Orkin, *Dana-Farber Cancer Institute, Boston, Massachusetts*

Panel Discussion: Historical and Prospective Glimpses into mRNA Splicing
Chairpersons: P. Sharp, *Massachusetts Institute of Technology, Cambridge;* J. Steitz, *HHMI/Yale University, New Haven, Connecticut*

Therapeutic Approaches to mRNA Splicing Diseases
Chairpersons: R. Kole, *Sarepta Therapeutics, Cambridge, Massachusetts;* C. Query, *Albert Einstein College of Medicine, Bronx, New York*

Genome Informatics

November 1–4 345 participants

ARRANGED BY **Janet Kelso,** Max Planck Institute, Leipzig, Germany
Aaron Quinlan, University of Utah, Salt Lake City
Melissa Wilson Sayres, Arizona State University, Tempe

This 15th CSHL/Wellcome Trust meeting continues to demonstrate a robust attendance, both from the United States and from abroad, attracting 345 registrants, presenting 230 abstracts, and offering a snapshot of the latest developments in the field. There were 12 invited talks and two keynote addresses. The remaining 36 talks were all selected for presentation by session chairs from openly submitted abstracts, and we also had 182 posters presented.

M. Wilson Sayres, A. Quinlan, J. Kelso

This year, abstracts covered a wide variety of genomic analyses, with a special emphasis on innovations in genetic variant discovery, data visualization, and new insights gained from integrating large-scale genomics data sets. There was also an emphasis on single-cell genome analysis, as well as gene isoform quantification and characterization. The sessions were centered around the topics listed in the Program below.

The first keynote address was delivered by Dr. Maricel Kann, who focused on strategies for interpreting tumor mutations based on mutational patterns observed in conserved protein domains. The second keynote was delivered by Dr. Lior Pachter, who discussed the consequences of using preprocessed experimental results instead of raw experimental data. Both talks were extremely well regarded and led to extensive follow-up discussions.

This meeting was very actively discussed on Twitter (using hashtag #gi2017), with the social media policy being opt-out (with virtually all talks being tweetable). More than 2000 messages were broadcast by users around the world to discuss and debate the ideas presented.

This meeting was funded in part by the National Human Genome Research Institute and the National Cancer Institute, branches of the National Institutes of Health; and Repositive.

M. Dubarry, V. Meyer

P. Ye, C. McLean

V. Schneider, P. Kitts

T. Hunt, J. Ling

PROGRAM

Variant Discovery and Genome Assembly
Chairpersons: L. Clarke, *EMBL-EBI, Cambridge, United Kingdom;* J. Simpson, *Ontario Institute for Cancer Research, Toronto, Canada*

Transcriptomics, Alternative Splicing, Gene Predictions
Chairpersons: M. Pertea, *Johns Hopkins University, Baltimore, Maryland;* O. Stegle, *EMBL-EBI, Cambridge, United Kingdom*

Keynote Speaker
M. Kann, *University of Maryland, College Park*

Data Curation and Visualization
Chairpersons: G. Marth, *University of Utah, Salt Lake City;* A. Loraine, *University of North Carolina, Charlotte*

Comparative and Metagenomics
Chairpersons: P. Flicek, *EMBL-EBI, Hinxton, United Kingdom;* H. Bik, *University of California, Riverside*

Epigenomics and Noncoding Genome
Chairpersons: E. Rivas, *Harvard University, Cambridge, Massachusetts*; A. Siepel, *Cold Spring Harbor Laboratory*

Keynote Speaker
L. Pachter, *University of California, Berkeley*

Personal and Medical Genomics
Chairpersons: K. Karczewski, *Broad Institute, Cambridge, Massachusetts;* S. Leal, *Baylor College of Medicine, Houston, Texas*

Single Cell Analyses

November 8–11 207 Participants

ARRANGED BY **Nancy Allbritton,** University of North Carolina, Chapel Hill
Scott Fraser, University of Southern California, Los Angeles
Junhyong Kim, University of Pennsylvania, Philadelphia

S. Fraser, N. Allbritton, J. Kim

The goal of this fifth workshop was to bring together scientists who analyze and engineer single cells using a wide variety of experimental paradigms to discuss the progress that is being made. More than 200 scientists convened with two keynote talks, 30 talks, and 75 posters. This year, a special session was organized around the theme of the evolutionary biology of single cells, which was seen to provide the theoretical foundations for studying both the origins and the proper classification of cell types. Participants in the meeting presented work addressing many different cell types ranging from bacteria to *Caenorhabditis elegans* and mammalian cells. The presentations were intentionally diverse in the goals of the studies and the range of techniques that are used to investigate single-cell biology, including single-cell dissection and transcriptomics, single-cell genomic DNA sequencing, proteomics, and metabolomics, all at single-cell resolution. Some talks concentrated on the dramatic recent advance in controlling single cells through engineered small-molecule receptors and photoresponsive proteins. A common theme from many of the talks and posters was individual cell-to-cell variability and how it relates to organismal function. Several talks discussed translational application of single-cell assays, and there was particular progress in the single-cell analysis of human cells discussed by several presenters. Several speakers presented novel molecular probes for individual cell function and methods to mechanically manipulate individual cell interactions. The importance of and necessity for microfabrication and nanotechnology were highlighted in several talks. There was continued effort at developing multimodal measurements at the single-cell level assisted

The Single Cell Analyses group

F. Quetier, N. Bolduc

P. Fordyce, N. Slovov

by these technologies. The need for quantitative analysis of biological properties/processes in live cells in their natural microenvironment and moving toward a comprehensive model of molecular cell function was a consistent theme. The results presented in the meeting offered ample evidence that the analysis of single-cell biology will lead to a better understanding of normal human health and disease states and better production of biofuels and insight into evolutionary mechanics, each of which would be unapproachable using pooled cell analyses.

This meeting was funded in part by Cell Microsystems; Protein Simple; Takara Bio; and 10x Genomics.

PROGRAM

Keynote Speaker
E. Rothenberg, *California Institute of Technology, Pasadena*

Probing the Limits of Measurement
Chairperson: S. Fraser, *University of Southern California, Los Angeles*

Dissecting the Single-Cell Genome
Chairperson: E. Rothenberg, *California Institute of Technology, Pasadena*

Flash Talks
Chairperson: N. Allbritton, *University of North Carolina, Chapel Hill*

Single-Cell Proteome and the Nucleome
Chairperson: N. Allbritton, *University of North Carolina, Chapel Hill*

Evolutionary Biology of the Cell
Chairperson: J. Kim, *University of Pennsylvania, Philadelphia*

Keynote Speaker
E.S. Boyden, *Media Lab and McGovern Institute, Massachusetts Institute of Technology, Cambridge*

L. Hayward, K. Kuritz

Complex Single-Cell Systems
Chairperson: J. Kim, *University of Pennsylvania, Philadelphia*

STATs: Importance in Basic and Clinical Cancer Research

November 15–18 84 Participants

ARRANGED BY **James Darnell,** The Rockefeller University, New York, New York
David Levy, New York University School of Medicine, New York
Valeria Poli, University of Turin, Italy
George Stark, Cleveland Clinic Lerner Research Institute, Ohio

V. Poli, D. Levy

This meeting, the first CSH meeting devoted to the biology of STAT proteins, grew out of a preliminary meeting at the Banbury Center organized to explore the role of STAT3 in human cancer. The themes identified at that meeting were expanded to explore the diverse roles of STAT family proteins in gene transcription, inflammation, immunity, metabolism, and differentiation, with an emphasis on how these functions impact oncogenesis. A common theme of many presentations revolved around approaches to manipulate STAT function, often through inhibition, in order to provide novel therapeutic approaches to human disease. An emerging theme was the diverse functions and subcellular organellar locations of individual STAT proteins, with particular interest in the multiple actions of STAT3 in the nucleus, mitochondria, lysosomes, and endoplasmic reticulum that impact metabolism, proton equilibria, calcium signaling, apoptosis, mitochondrial protein translation, and cell identity. Another theme was the genetic STAT-dependent diseases, including novel information concerning the mechanisms underlying inherited immunodeficiency and autoinflammatory diseases, the roles of activated kinases in malignancy, and the signaling networks that regulate STAT-dependent transcriptional function. A major interest of many participants was the development and analysis of small-molecule inhibitors of STAT-dependent processes that could have therapeutic efficacy. A number of important developments in this area were discussed, including novel approaches to the design and analysis of

The STATs group

Y. Li, C. Watson

O. Danziger, M. Rincon

inhibitor action, approaches to improved specificity and affinity, and the distinction between inhibition of upstream activators and direct inhibition of STAT protein function. Another important concept discussed was the expansion of the paradigm of JAK-STAT signaling as a pathway governed by STAT protein tyrosine phosphorylation to include a number of STAT-dependent processes that are independent of tyrosine phosphorylation, being dependent instead on alternative modification, protein abundance, or protein–protein interactions. The results presented at the meeting highlighted the many molecular mechanisms that are common across the STAT protein family, the important points of intersection between different family members, and the multiple biological systems that are regulated by STAT proteins and interact to produce diverse physiologic and pathophysiologic outcomes. The presentations at the meeting made abundantly clear that numerous STAT-dependent processes and mechanisms remain to be documented and elucidated, including additional areas that will likely provide additional paradigm-shifting insights.

PROGRAM

JAK-STAT Mutations and Genomic Functions
Chairperson: J. Darnell, *The Rockefeller University, New York, New York*

JAK-STAT Signaling in the Hematopoietic System
Chairpersons: C. Mertens, *The Rockefeller University, New York, New York*; D. Levy, *New York University Langone Medical Center, New York, New York*

STAT3, Metabolism, and Cancer
Chairperson: R. Jove, *Nova Southeastern University, Fort Lauderdale, Florida*

JAK-STAT Signaling and Regulation
Chairpersons: V. Poli, *University of Turin, Italy;* G. Stark, *Cleveland Clinic Lerner Research Institute, Ohio*

B. Maurer, W. Leonard

R. Jove, J. Darnell

U. Vinkemeier, G. Stark

Transcriptional and Epigenetic Regulation by STATs
Chairperson: I. Marie, *New York University School of Medicine, New York*

JAK-STAT Inhibition
Chairpersons: J. Turkson, *University of Hawaii Cancer Center, Honolulu;* J. Darnell, *The Rockefeller University, New York, New York*

Plant Genomes and Biotechnology: From Genes to Networks

November 29–December 2 132 Participants

ARRANGED BY **David Jackson,** Cold Spring Harbor Laboratory
Todd Mockler, Donald Danforth Plants Science Center, Saint Louis, Missouri
Jane Parker, Max Planck Institute for Plant Breeding Research, Münich, Germany
Seung (Sue) Rhee, Carnegie Institution for Science, Washington, D.C.

This 11th plants meeting demonstrated the power of genome-enabled plant biology in a broad spectrum of areas ranging from environmental adaptation to developmental network modeling and crop improvement. Participants and presenters were a healthy mix of junior and established scientists, staying true to the vision of the first meeting in 1997 and the spirit of Cold Spring Harbor Laboratory. A breadth and depth of exciting, new, unpublished work toward discovery of fundamental principles and direct applications in agriculture was presented.

T. Mockler, D. Jackson, J. Parker

A major theme across sessions was disruptive tools being developed or applied to address technical challenges and answer key biological questions. For example, some talks discussed the genomic architecture of important crop traits and identification of genes underlying these traits that will be useful for breeders. Other talks introduced bioengineering approaches to piece together the mechanical, spatial, and temporal dynamics of important crop traits. Also shown were rapid advances in protein structural biology, which are impacting the design of protein functional modules for modulating development and pathogen recognition.

Another major theme was disruptive concepts and out-of-the-box thinking. Examples included work aimed at developing new, safe chemical treatments to alleviate crop losses from drought, research on global identification of transcriptional regulatory elements in plants, and efforts to assign


The Plant Genomes and Biotechnology group

A. Eveland, R. Parvathaneni

M. Gore, P. Qiao, M. Kramer

molecular functions to the "dark matter" of genomes, which remains a problem even for the most intensively studied model organisms. Another disruptive technology is CRISPR-Cas9, and advances in developing this method for genome editing in *Arabidopsis* and crop plants were presented.

Another prominent thread in the meeting related to the rapid development and adoption of high-throughput phenotyping in model and crop plants. These new technologies and approaches are enabling the characterization of developmental processes at many scales and uncovering of genetic variation underlying potentially useful traits. Other reports presented new insights into genome-wide analysis of primary and secondary metabolic pathways with the promise to engineer new beneficial plant products. These and other studies described at the meeting emphasize the multidisciplinary nature of genome-enabled quantitative biology to harness new phenotypic variation and genetic epistasis for crop improvement. Descriptions of state-of-the-art, high-throughput technologies and genome biology were complemented nicely by presentations on genetic, molecular, and spatial dissections of biological processes in model and crop species, including important new insights into plant defense mechanisms.

The keynote address was presented by Dr. Julia Bailey-Serres, who focused on the integration of omics approaches to understand plant responses to water stress, relevant to improving global production of rice and other staple crops.

PROGRAM

Keynote Speaker

J. Bailey-Serres, *University of California Riverside Center for Plant Cell Biology, Riverside/Utrecht University, Netherlands*

Frontier Technologies and Synthetic Biology

Chairperson: S. Navlakha, *The Salk Institute for Biological Studies, La Jolla, California*

E. Buckler, A. Kent

Z. Brenton, J. Bailey-Serres

Crop Biology and Trait Enhancement
Chairperson: M. Paul, *Rothamsted Research, Harpenden, United Kingdom*

Metabolic Circuits
Chairperson: A. Osbourn, *John Innes Centre, Norwich, United Kingdom*

Plants and Microbes
Chairperson: M. Harrison, *Boyce Thompson Institute, Ithaca, New York*

Development: Modules to Networks
Chairperson: T. Ito, *Nara Institute of Science and Technology, Japan*

Biodiversity and Environmental Adaptation
Chairperson: J Bergelson, *University of Chicago, Illinois*

Keynote Speaker
R. Martienssen, *Cold Spring Harbor Laboratory*

Genomes and Epigenomes
Chairperson: D. Voytas, *University of Minnesota, St. Paul*

Development and 3D Modeling of the Human Brain

December 6–9 118 Participants

ARRANGED BY **Paola Arlotta,** Harvard University, Cambridge, Massachusetts
Sergiu Pasca, Stanford University, California

The goal of this inaugural meeting was to bring together leaders in the field of human brain development, evolution, stem cell biology, and neuropsychiatric disease to discuss progress in these fields and how they are informing 3D modeling of the human brain using human brain organoids. Participants convened with two keynote talks, 33 talks, and 42 posters. The presentations were intentionally diverse in the goals of the studies and the range of techniques that are used in this emerging field. The meeting started with a keynote on hominid evolution covering breakthroughs from deciphering ancient genomes and approaches on how these, in combination with human cellular models, may inform uniquely human phenotypes. The first session covered recent advances in developmental human neurobiology and introduced work on organoids as tools for studying cellular diversity in the nervous system. Then, participants in the meeting presented various 3D models of the human brain applied to studying oligodendrogenesis, Zika-related microcephaly, tumorigenesis, and various genetically defined neurodevelopmental and neurodegenerative disorders, as well as approaches for assembling brain regions in vitro to study circuit formation. The second keynote addressed the ethical concerns of human brain surrogate models and ways these could be addressed by engaging the scientific community and the public. The importance and necessity of highly parallel single-cell gene expression profiling and of establishing robust assays in 3D cultures (including electrophysiological methods) was highlighted in several talks. Multiple speakers presented new technologies that could have applications for studying brain organoids, including CRISPR-Cas9 screens and a range of bioengineering techniques.

The Brain group

N. Sestan, A.R. Kriegstein

Z-b. Tong, S. Szuchet

PROGRAM

Keynote Speaker
S. Pääbo, *Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany*

Human Brain Development and Evolution I
Chairperson: H. Song, *University of Pennsylvania, Philadelphia*

Human Brain Development II
Chairperson: F. Vaccarino, *Yale University, New Haven, Connecticut*

Keynote Speaker
I. Hyun, *Case Western Reserve University School of Medicine, Cleveland, Ohio*

Talks from Abstracts
Chairperson: S. Pasca, *Stanford University, California*

Brain Models, Development, and Disease I
Chairperson: K. Eggan, *Harvard University, Cambridge, Massachusetts*

Brain Models, Development, and Disease II
Chairperson: J.A. Knoblich, *IMBA, Austrian Academy of Sciences, Vienna*

Brain Models, Development, and Disease III
Chairperson: G-l. Ming, *University of Pennsylvania, Philadelphia*

POSTGRADUATE COURSES

High-Throughput Biology: From Sequence to Networks

March 20–26

INSTRUCTORS **A. Meyer**, Ontario Institute for Cancer Research, Toronto, Ontario, Canada
F. Ouellette, Genome Quebec, Montreal, Canada

TEACHING ASSISTANTS **Z. Lu,** Princess Margaret Cancer Centre, Toronto, Ontario, Canada

With the introduction of next-generation sequencing platforms, it is now feasible to use high-throughput sequencing approaches to address many research questions. Now more than ever, it is crucial to know what bioinformatics tools and resources are available, and it is necessary to develop informatics skills to analyze high-throughput data using those tools. The Canadian Bioinformatics Workshops (CBW), in collaboration with Cold Spring Harbor Laboratory, has developed a comprehensive seven-day course covering key bioinformatics concepts and tools required to analyze DNA and RNA sequence reads using a reference genome. This course combined the material and concepts from three established CBW workshops.

The course began with the workflow involved in moving from platform images to sequence generation, after which participants gained practical skills for evaluating sequence read quality, mapped reads to a reference genome, and analyzed sequence reads for variation and expression level. The course concluded with pathway and network analysis on the resultant "gene" list. Participants gained experience in cloud computing and data visualization tools. All class exercises were self-contained units that included example data (e.g., Illumina paired-end data) and detailed instructions for installing all required bioinformatics tools.

This course was supported with funds provided by the Helmsley Charitable Trust and the Howard Hughes Medical Institute and Computational Resources supported by Amazon Web Services.

PARTICIPANTS

Alpaugh, W., Ph.D., University of Calgary, Alberta, Canada
Beecham, A., M.S., University of Miami, Coral Gables, Florida
Bosch, A., Ph.D., Mount Sinai School of Medicine, New York
Cook, E., M.D./Ph.D., Queen's University, Kingston, Ontario, Canada
Cox, B., Ph.D., University of Toronto, Ontario, Canada
Doan, T., Ph.D., University of California, San Francisco
Domingo, A., Ph.D., Massachusetts General Hospital, Boston
Edwards, D., Ph.D., Icahn School of Medicine at Mount Sinai, New York
El Rouby, N., Ph.D., University of Florida, Gainesville
Gonzalez-Buendia, E., Ph.D., Columbia University, New York, New York
Karim, R., B.S., University of Cincinnati, Ohio
Kennedy, A., Ph.D., Guardant Health, Inc., Redwood City, California
Kodysh, J., M.Eng., Icahn School of Medicine, Mount Sinai, New York
Lazaris, C., M.Sc., New York University School of Medicine, New York
Lowther, C., Ph.D., University of Toronto, Ontario, Canada
Mamiya, P., Ph.D., University of Washington, Seattle
Muhle, R., M.D./Ph.D., Yale University, New Haven, Connecticut
Pantelireis, N., M.Eng., Imperial College London, United Kingdom
Patterson, J., Ph.D., Michigan State University, Grand Rapids
Revuelta, M., Ph.D., Weill Cornell Medical College, New York, New York
Schönherz, A., Ph.D., Aalborg University, Denmark
Song, X., B.S., Baylor College of Medicine, Houston, Texas
Stevens, S., Ph.D., University of South Florida, Tampa
Stiller, B.M., Ph.D, German Heart Center, Technical University Münich, Germany
Tutaj, M., Ph.D., Medical College of Wisconsin, Milwaukee
Wang, J., B.S., Baylor College of Medicine, Houston, Texas
Zavallo, D., Ph.D., National Institute of Agricultural Technology, Hurlingham, Argentina
Zhao, X., Ph.D., Moffitt Cancer Center, Tampa, Florida

SEMINARS

Simpson, J., Ontario Institute for Cancer Research, Toronto, Canada: Module 1: Introduction to HT sequencing.
Cavalli, F., University of Toronto, Canada: Module 2: Data visualization.
Bourgey, M., McGill University, Montreal, Quebec, Canada: Module 3: Genome alignment.
Lu, Z., Princess Margaret Cancer Centre, Toronto, Ontario, Canada: Module 3: Connecting to the Cloud.
Bourgey, M., McGill University, Montreal, Quebec, Canada: Module 4: Small-variant calling and annotation. Module 5: Structural variant calling.
Simpson, J., Ontario Institute for Cancer Research, Toronto, Canada: Module 6: De novo assembly.
Griffith, M., Washington University School of Medicine in St. Louis, Missouri: Module 7: Introduction to RNA sequencing and analysis.
Yousif, F., The Ontario Institute for Cancer Research, Toronto, Canada: Module 8: RNA-Seq alignment and visualization.
Griffith, O., Washington University School of Medicine in St. Louis, Missouri: Module 9: Expression and differential expression.
Griffith, M., Washington University School of Medicine in St. Louis, Missouri: Module 10: Reference-free alignment. Module 11: Isoform discovery and alternative expression.
Reimand, J., Ontario Institute for Cancer Research, Toronto, Canada: Module 12: Introduction to pathway and network analysis. Module 13: Finding overrepresented pathways. Module 13 Lab: Enrichment-based analysis: Performing ORA.
Voisin, V., University of Toronto, Ontario, Canada: Module 14: Cytoscape intro, demo, enrichment map. Module 14 Lab: Cytoscape intro, demo, enrichment map.
Haw, R., Ontario Institute for Cancer Research, Toronto, Canada: Module 15: More depth on network and pathway and analysis. Module 15 Lab: De novo subnetwork clustering analysis: Reactome.
Morris, Q., University of Toronto, Toronto, Ontario, Canada: Module 16: Gene function prediction.
Voisin, V., University of Toronto, Ontario, Canada: Module 16 Lab: Gene-mania.
Hoffman, M., Princess Margaret Cancer Centre, Toronto, Ontario, Canada: Module 17: Gene regulation network analysis.
Voisin, V., University of Toronto, Ontario, Canada: Module 17 Lab: iRegulon.

Workshop on Leadership in Bioscience

March 24–27

INSTRUCTORS **C. Cohen,** Science Management Associates, Newton, Massachusetts
S. Cohen, Science Management Associates, Newton, Massachusetts

In this highly interactive 3.5-day workshop, students developed the skills necessary to lead and interact effectively with others in both one-on-one and group settings. The workshop focused on techniques, situations, and challenges that relate specifically to leading and managing in the scientific workplace. It emphasized learning by doing and involved role playing, giving and receiving feedback, and group problem solving. Much of the learning was peer-to-peer. Participants were expected to discuss their own experiences and listen to others as they discuss theirs.

The workshop helped participants identify areas where they need guidance and growth, as well as how to capitalize on areas of strength. They had the opportunity to share their experiences and challenges with others and to receive feedback and guidance from others with experience in leading scientists in a variety of settings. At the end of the course, participants were linked through a unique online community in which they continued to learn from one another and from the course instructors. Key focus areas of the workshop included:

- Recognizing and understanding leadership in a science setting.
- Using negotiation as a tool in scientific discussions and problem solving.
- Identifying and resolving conflicts in the lab.
- Dealing with difficult people and situations.
- Communicating your ideas and plans in a way that engages others.
- Leading effective and productive meetings.
- Becoming effective citizen scientists.

The workshop was targeted to life scientists making, or recently having made, the transition to leadership or managerial positions. Many of the situations discussed were from the perspective of independent investigators running their own laboratory. As such, relatively new investigators (e.g., <3 years) were particularly encouraged to apply, as were senior postdoctoral scholars on the cusp of tenure-track research positions.

This course was supported with funds provided by National Institute of General Medical Sciences.

PARTICIPANTS

Abi-Ghanem, D., Ph.D., Immunology Consultants Laboratory, Portland, Texas
Barr, J., Ph.D., Monash University, Clayton, Victoria, Australia
Braun, J., Ph.D., University of Massachusetts Medical School, Worcester
Connelly, C., Ph.D., Streck, LaVista, Nebraska
DeBlasio, S., Ph.D., USDA-ARS, Ithaca, New York
Depis, F., Ph.D., Merrimack Pharmaceuticals, Cambridge, Massachusetts
Egekwu, N., Ph.D., ARS-USDA, Beltsville, Maryland
Ehmsen, J., M.D./Ph.D., Johns Hopkins School of Medicine, Baltimore, Maryland
Engle, D., Ph.D., Cold Spring Harbor Laboratory
Golkar, Z., Ph.D., Voorhees College, Denmark, South Carolina
Greschner, A., Ph.D., Institut National de la Recherche Scientifique, Varennes, Quebec, Canada
Kim, Y., Ph.D., Sungkyunkwan University, Seongnam, South Korea
Kirby, E., Ph.D., The Ohio State University, Columbus
Kroeger, Jr., P., Ph.D., University of Pennsylvania, Philadelphia
Lauberth, S., B.S., University of California, San Diego
Lewis, T., Ph.D., Columbia University, New York, New York
Mayhew, M., Ph.D., Lawrence Livermore National Laboratory, Livermore, California
McHugh, C., Ph.D., California Institute of Technology, Pasadena
Nityanandam, A., Ph.D., St. Jude Children's Research Hospital, Memphis, Tennessee
Oudin, M., Ph.D., Massachusetts Institute of Technology, Cambridge
Ozcan, L., M.D., Columbia University, New York, New York
Reimand, J., Ph.D., Ontario Institute for Cancer Research, Toronto, Canada
Schröder, B., Ph.D., Gothenburg University, Sweden
Wilen, C., M.D./Ph.D., Washington University School of Medicine in St. Louis, Missouri

SEMINARS

Cohen, C., and Cohen, S., Science Management Associates, Newton, Massachusetts: Session 1: Who We Are. Session 2: Leadership Challenges: Case Study Overview. Session 3: Difficult Conversations and Interactions: Fundamentals of Negotiation. Session 4: Case Study Analysis, Part 1.
Qualls, M., Entropia Consulting, Philadelphia, Pennsylvania: Session 5: The Impact of Your Leadership Style on Team Management.
Cohen, C., and Cohen, S., Science Management Associates, Newton, Massachusetts: Session 6: Scientific Team and Project Meetings. Session 7: Projecting Leadership, Part 1. Session 8: Case Study Analysis, Part 2. Session 9: Managing Your Science Team: Goal Setting, Feedback, and Evaluation for Scientist. Session 10: Concluding Group Discussion

Cell and Developmental Biology of *Xenopus*

March 29–April 11

INSTRUCTORS **M. Khokha,** Yale University, New Haven, Connecticut
K. Liu, King's College London, United Kingdom

ASSISTANTS **P. Date,** Yale University, New Haven, Connecticut
R. Huebner, The University of Texas, Austin
E. Mis, Yale University, New Haven, Connecticut
H. Rankin Willsey, University of California, San Francisco

Xenopus is remarkable for modeling human diseases including birth defects, cancer, and stem cell biology. *Xenopus* continues to make a major impact on our understanding of cell and developmental biology. Students were encouraged to target genes of interest using CRISPR technology and then analyzed phenotypes using the diverse array of assays available in *Xenopus*. Specifically, techniques covered included microinjection and various molecular manipulations including CRISPR knockouts, morpholino-based depletions, transgenics, and mRNA overexpression. In addition, students combined these techniques with explant and transplant methods to simplify or test tissue-level interactions. To visualize subcellular and intercellular activities, we introduced a variety of imaging methods including time-lapse, fluorescent, and confocal microscopy. Additional methods included mRNA in situ hybridization and protein immunohistochemistry as well as basic bioinformatic techniques for gene comparison and functional analysis. Biochemical approaches such as proteomics and mass spectrometry were also discussed. This course was designed for those new to the *Xenopus* field, as well as for more advanced students who were interested in emerging technologies. We encouraged students to bring their own genes of interest and to tailor aspects of the course to enable them to initiate studies on their specific projects.

This course was supported with funds provided by National Institute of Child Health and Human Development.

PARTICIPANTS

Alkobtawi, M., M.A., Paris-Saclay University, Orsay, France
Cádiz-Rivera, B., M.S., University of Puerto Rico, Humacao
Golding, A., B.S., University of Wisconsin, Madison
Jarvis Alberstat, E., M.S., University of California, Berkeley
Kyung Song, E., B.S., Ulsan National Institute of Science and Technology, South Korea
Lasser, M., B.S., Boston College, Chestnut Hill, Massachusetts
Ng, N., B.Sc., University of Manchester, United Kingdom
Portero, E., B.A., The George Washington University, Washington, D.C.
Ray, H., M.S., University of Alabama, Birmingham
Reeve, R., M.S., Washington State University, Pullman
Scott, C., B.Sc., University of Manchester, United Kingdom
Stuckenholz, C., Ph.D., University of Pittsburgh, Pennsylvania

SEMINARS

Blitz, I., University of California, Irvine: Leapfrogging to fast track making knockout lines.
Cha, S-W., Cincinnati Children's Hospital Medical Center, Ohio: HDR-mediated precision genome editing in *Xenopus*.
Conlon, F., University of North Carolina, Chapel Hill: *Xenopus* proteomic approaches to development and disease.
Davidson, L., University of Pittsburgh, Pennsylvania: Leveraging *Xenopus* to explore the mechanics and mechanobiology of development.
Mayor, R., University College London, United Kingdom: Neural crest migration.
Heald, R., University of California, Berkeley: Mechanisms of mitosis and size control in *Xenopus.*
Keller, R., University of Virginia, Charlottesville: Early *Xenopus* morphogenesis.
Liu, K., King's College London, United Kingdom: *Xenopus*: Not just a pretty face.
Khokha, M., Yale University, New Haven, Connecticut: Patient-based gene discovery and *Xenopus*-based mechanism discovery.
Miller, A., University of Michigan, Ann Arbor: Using *Xenopus* to investigate epithelial cytokinesis and cell–cell junction dynamics.
Nemes, P., George Washington University, Washington D.C.: Mass spectrometry for cell/developmental biology of *Xenopus*.
Thomsen, G., Stony Brook University, New York: *Xenopus:* Show and tell demo.
Steventon, B., University of Cambridge, United Kingdom: Image processing.
Wallingford, J., University of Texas, Austin: The majesty of the frog: The past, present, and future of research with *Xenopus*.
Wills, A., University of Washington, Seattle; Chromatin and nuclear remodeling in *Xenopus* tail regeneration.
Witkowski, J., Cold Spring Harbor Laboratory: Ethics, reproducibility, and vigor.

Expression, Purification, and Analysis of Proteins and Protein Complexes

March 29–April 11

INSTRUCTORS
A. Courey, University of California, Los Angeles
S-H. Lin, MD Anderson Cancer Center/University of Texas, Houston
M. Marr, Brandeis University, Waltham, Massachusetts
S. Nechaev, University of North Dakota School of Medicine, Grand Forks

ASSISTANTS
A. Bhattacharia, University of North Dakota, Grand Forks
N. Clark, Brandeis University, Waltham, Massachusetts
S. Ghosh Dastidar, University of North Dakota, Grand Forks
Y-C. Lee, MD Anderson Cancer Center, Houston, Texas
A. Sawyer, Brandeis University, Waltham, Massachusetts
T. Yau, University of California, Los Angeles

This course was intended for scientists, including graduate students, postdoctoral scholars, staff scientists, and principal investigators, who wanted a rigorous introduction to expression and purification of proteins as well as analysis of protein structure and function.

Through hands-on experience in the lab as well as extensive lecture and discussion, each student became familiar with key approaches in expression, purification, and analysis of soluble and membrane proteins and protein complexes from both natural sources and overexpression systems. The emphasis of the course was on the following:

1. Approaches in protein expression: Choosing the best bacterial or eukaryotic expression system tailored for the particular protein and experimental problem; Determining how to optimize expression; Understanding protein tagging: The advantages and pitfalls of various affinity and solubility tags.

2. Approaches in protein purification: Choosing the best strategy for a given protein, including solubilization; Bulk fractionation; Liquid chromatography, including conventional methods (ion exchange, size exclusion, reverse phase, etc.) and affinity methods (e.g., MAC, DNA affinity, immunoaffinity, etc.), as well as FPLC/HPLC.
3. Approaches in protein analysis: Introduction to common approaches for characterization of proteins, including binding assays; Activity assays; Mass spectroscopy to identify protein interaction partners and posttranslational modifications.

In addition to purification, students gained exposure to fundamental analytical approaches such as mass spectroscopy and protein structure determination (e.g., X-ray crystallography and cryo-EM).

PARTICIPANTS

Arnling Bååth, J., M.S., Chalmers University of Technology, Gothenburg, Sweden
Bastle, R., M.A., Icahn School of Medicine Mount Sinai, New York
Bhattacharjee, S., Ph.D., Cold Spring Harbor Laboratory
Cantoria, M., Ph.D., University of Texas Southwestern Medical Center, Dallas
Dornan, M., Ph.D., University of Montreal, Quebec, Canada
Ford, L., Ph.D., Columbia University, New York, New York
Gamero Estevez, E., B.Sc., McGill University, Montreal, Quebec, Canada
Halova, L., Ph.D., University of Manchester, United Kingdom
Kernan, L., B.A., University of North Carolina, Chapel Hill
Kurten, C., M.S., Royal Institute of Technology, Solna, Sweden
Li, S., Ph.D., Cold Spring Harbor Laboratory
Martinez, J., Ph.D., University of Delaware, Newark
Sachan, N., B.S., Cold Spring Harbor Laboratory
Sarr, M., M.S., Karolinska Institutet, Huddinge, Sweden
Shan, X., B.S., Stony Brook University, New York
Subramanian, N., Ph.D., University of Calgary, Canada

SEMINARS

Burgess, R., University of Wisconsin, Madison: Introduction to protein purification/refolding insoluble proteins. Gentle immunoaffinity chromatography and a new technique for identifying and studying weak protein–protein interactions.
Courey, A., University of California, Los Angeles: Proteomic analysis of Groucho and SUMO partners in *Drosophila* embryogenesis.
Jiang, J., University of California, Los Angeles: High-resolution cryo-EM single-particle reconstruction of protein complexes.
Love, J., Expression Technologies, Newark, California: Expression and purification of mammalian proteins.
Lin, S-H., MD Anderson Cancer Center/University of Texas, Houston: Prostate cancer bone metastasis.
Marr, M., Brandeis University, Waltham, Massachusetts: Gene regulation in response to cellular stress.
Nechaev, S., University of North Dakota School of Medicine, Grand Forks: Responses to stimuli may involve activation of promoters without activation of their genes.
Pappin, D., Cold Spring Harbor Laboratory: Introduction to mass spectrometry of proteins. Quantitative approaches to mass spectrometry of proteins.
Witkowski, J., Cold Spring Harbor Laboratory: Ethics, reproducibility, and vigor.

Quantitative Imaging: From Acquisition to Analysis

March 29–April 11

INSTRUCTORS **H. Elliott,** Harvard Medical School, Boston, Massachusetts
J. Waters, Harvard Medical School, Boston, Massachusetts

CO-INSTRUCTORS **T. Lambert,** Harvard Medical School, Boston, Massachusetts
D. Richmond, Harvard Medical School, Boston, Massachusetts

TEACHING ASSISTANTS **M. Cicconet,** Harvard Medical School, Boston, Massachusetts
A. Jost, Harvard Medical School, Boston, Massachusetts
P. Montero Llopis, Harvard Medical School, Boston, Massachusetts
M. Weber, Harvard Medical School, Boston, Massachusetts

Combining careful image acquisition with rigorous computational analysis allows extraction of quantitative data from light microscopy images that is far more informative and reproducible than what can be seen by eye. This course focused on advanced quantitative fluorescence microscopy techniques used for imaging a range of biological specimens, from tissues to cells to single molecules. The course was designed for quantitative cell and molecular biologists, biophysicists, and bioengineers.

We provided a thorough treatment of the complete process of quantitative imaging, from the photons emitted from the sample to the extraction of biologically meaningful measurements from digital images. Material was covered in lectures, discussion groups, and hands-on quantitative exercises using commercial microscopes and open-source image analysis tools. The concepts covered included wide-field fluorescence microscopy, laser scanning and spinning-disk confocal microscopy, CCD, EM-CCD, and sCMOS cameras, total internal fluorescence microscopy (TIRF), light-sheet microscopy, super-resolution microscopy (structured illumination, STED, and localization microscopy), imaging and analyzing ratiometric "biosensors" (including FRET), fluorescent proteins and live-sample imaging, image processing (filtering, de-noising, corrections, and deconvolution),

image segmentation, quantitative shape and intensity measurements, object detection and tracking, machine learning, and designing and troubleshooting quantitative imaging experiments.

The course also included a series of seminars from guest speakers who applied the methods we discussed.

This course was supported with funds provided by the Helmsley Charitable Trust and the Howard Hughes Medical Institute and Major support provided by the National Cancer Institute.

PARTICIPANTS

Chang-Graham, A., B.A., Baylor College of Medicine, Houston, Texas
Dasgupta, A., Ph.D., University of Texas Southwestern Medical Center, Dallas
Deo, C., Ph.D., Howard Hughes Medical Institute, Ashburn, Virginia
Dobramysl, U., Ph.D., University of Cambridge, United Kingdom
Estrem, C., B.A., University of Colorado Anschutz Medical Campus, Aurora
Gambe-Gilbuena, A., Ph.D., Ateneo de Manila University, Quezon City, Philippines
Hanna, D., B.A., Georgia Institute of Technology, Atlanta
Leoncini, E., Ph.D., Harvard University, Boston, Massachusetts
Mino, R., Ph.D., University of Texas Southwestern Medical Center, Dallas
Monteiro, D., M.S., University of California, San Francisco
Mui, K., Ph.D., Columbia University, New York
Reilly, N., B.S., University of Rochester, New York
Rosselli, L., Ph.D., University of Michigan, Ann Arbor
Sales, E., B.S., University of Oregon, Eugene
Segal, G., Ph.D., The University of Melbourne, Victoria, Australia
Stolfi, A., Ph.D., New York University, New York

SEMINARS

Waters, J., Harvard Medical School, Boston, Massachusetts: Quantitative microscopy basics. Objective lenses. Transmitted light microscopy. Fluorescence microscopy.
Lambert, T., Harvard Medical School, Boston, Massachusetts: Digital imaging.
Elliott, H., Harvard Medical School, Boston, Massachusetts: Basics of image processing and digital microscopy: Resolution, SNR, and diffraction-limited objects.
Waters, J., Harvard Medical School, Boston, Massachusetts: Quantifying fluorescence: Image acquisition and controls.
Elliott, H., Harvard Medical School, Boston, Massachusetts: Image processing 2: Image corrections and advanced filtering.
Shaner, N., The Scintillon Institute, San Diego, California: Fluorescent proteins.
Elliott, H., Harvard Medical School, Boston, Massachusetts: Image segmentation and morphometry.
Waters, J., Harvard Medical School, Boston, Massachusetts: Live-cell imaging.
Elliott, H., Harvard Medical School, Boston, Massachusetts: Image correlation methods: Colocalization, registration, and stitching.
Waters, J., Harvard Medical School, Boston, Massachusetts: TIRF.
Lambert, T., Harvard Medical School, Boston, Massachusetts: Confocal microscopy theory and hardware.
Elliott, H., Harvard Medical School, Boston, Massachusetts: 3D image analysis and deconvolution.
Waters, J., Harvard Medical School, Boston, Massachusetts: Limitations on quantitative imaging of thick samples. Live confocal microscopy and intensity measurements over time.
Abrahamsson, S., The Rockefeller University, New York, New York: Multifocus microscopy in fast volumetric functional neuronal imaging.
Witkowski, J., Cold Spring Harbor Laboratory: Ethics lecture.
Elliott, H., Harvard Medical School, Boston, Massachusetts: Image time series analysis: Tracking, photobleach correction, and FRAP analysis.
Waters, J., Harvard Medical School, Boston, Massachusetts: Multiphoton microscopy.
Richmond, D., Harvard Medical School, Boston, Massachusetts: Machine learning in bioimage analysis.
Waters, J., Harvard Medical School, Boston, Massachusetts: Light sheet microscopy, Part 1.
Lambert, T., Harvard Medical School, Boston, Massachusetts: Light sheet microscopy, Part 2.
Fowlkes, C., University of California, Irvine: Learning to detect, segment, and track biological structures.
Huang, B., University of California, San Francisco: Super-resolution microscopy I localization. Super-resolution microscopy II patterned illumination.
Huisken, J., Morgridge Institute for Research, Madison, Wisconsin: Multiscale imaging with personalized instrumentation.
Manley, S., Ecole Polytechnique Federale de Lausanne (EPFL), Switzerland: Expanding horizons with large field-of-view, automated localization microscopy.

Advanced Bacterial Genetics

June 6–26

INSTRUCTORS **L. Bossi,** Institut de Biologie, Intégrative de la Cellule-12BC, Paris, France
A. Camilli, Tufts University Medical School, Boston, Massachusetts
A. Grundling, Imperial College London, United Kingdom

ASSISTANTS **R. Balbontin Soria,** Instituti Gulbekian de Ciencia, Oeiras, Portugal
N. Figueroa-Bossi, Institut de Biologie Integrative de la Cellule-I2BC, France
C. Schuster, Imperial College London, United Kingdom
L. Shull, Tufts University Medical School, Boston, Massachusetts

This course presented logic and methods used in the genetic dissection of complex biological processes in diverse bacteria. Laboratory methods included classical and cutting-edge mutagenesis using transposons, allelic exchange, and TargeTron; recombineering with single- and double-stranded DNA; CRISPR-Cas genome editing; genome sequencing and assembly; mapping mutations using genetic and physical techniques; modern approaches to the generation and analysis of targeted gene disruptions and reporter gene fusions; and fluorescence microscopy. Key components of the course were the use of sophisticated genetic methods in the analysis of model bacteria (including *Escherichia coli, Salmonella, Staphylococcus aureus,* and *Vibrio cholerae*) and the use of the wealth of new genomic sequence information to motivate these methods.

Invited lecturers presented various genetic approaches to study bacterial mechanisms of metabolism, development, and pathogenesis.

This course was supported with funds provided by the National Science Foundation.

PARTICIPANTS

Beblawy, S., Karlsruhe Institute of Technology, Germany
Blakeway, L., Ph.D., Griffith University, Gold Coast, Australia
Cooper, K., Ph.D., NIH/NIAID/Rocky Mountain Laboratories, Hamilton, Montana

Datta, M., Ph.D., Massachusetts Institute of Technology, Cambridge
Guillemette, R., Ph.D., University of California, San Diego
Klein, J., B.S., Washington State University, Pullman
Lynch, J., Ph.D., University of Hawaii, Manoa, Honolulu
Mullane, K., B.S., Scripps Institution of Oceanography, University of California, San Diego
Niketic, D., Ph.D., University of Utah, Salt Lake City
Raible, K., B.A., Drexel College of Medicine, Philadelphia, Pennsylvania
Rajagopalan, K., M.S., Columbia University, New York, New York
Szymczak, P., M.S., University of Copenhagen, Hoersholm, Denmark
Tan, J., Ph.D., Virginia Commonwealth University, Richmond
Tovar, K., B.A., University of Illinois, Chicago,
Vang Nielsen, S., B.S., University of Copenhagen, Denmark
Xue, B., Ph.D., Institute for Advanced Study, Princeton, New Jersey

SEMINARS

Boucher, Y., University of Alberta, Edmonton, Canada: Bacterial epidemiology in the era of (meta) genomics.
Casadesús, J., Universidad de Seville, Spain: Luria and Delbruck revisited: Nonmutational preadaptation to lethal selection.
LeMieux, J., Tufts University Medical School, Boston, Massachusetts: Science communication: Importance of scientists communicating and tips on how to more effectively communicate.
Marraffini, L., The Rockefeller University, New York, New York: CRISPR-Cas: The adaptive immune system of prokaryotes.
Pukatzki, S., University of Colorado School of Medicine, Aurora: The hunt for conserved virulence traits.
Segall, A., San Diego State University, California: Illuminating viral dark matter.
Wang, J., University of Wisconsin, Madison: Replication–transcription conflict in bacteria.
Young, R., Texas A&M University, College Station: Phage lysis: Do we have the whole story now?

Ion Channels in Synaptic and Neural Circuit Physiology

June 6–26

INSTRUCTORS **T. Branco,** Sainsbury Wellcome Center, London, United Kingdom
I. Duguid, University of Edinburgh, Midlothian, United Kingdom
C. Schmidt-Hieber, Pasteur Institute, Paris, France

CO-INSTRUCTORS **N. Wanaverbecq,** Aix-Marseille University, Marseille, France

ASSISTANTS **C. Arlt,** Harvard Medical School, Boston, Massachusetts
K. Betsios, Mantis64, Athens, Greece
L. Brosse, Aix-Marseille University, Marseille, France
E. Galliano, Harvard University, Cambridge, Massachusetts
J. McCauley, University at Albany, Queensbury, New York

Ion channels are the fundamental building blocks of excitability in the nervous system. The primary goal of this course was to demonstrate, through lectures and laboratory work, the different biophysical properties of ion channels that enable neurons to perform unique physiological functions in a variety of neural systems.

Areas of particular interest included (1) voltage- and ligand-gated ion channels at central and peripheral synapses, (2) synaptic integration and plasticity, (3) neural circuit function in vitro and in vivo, and (4) optogenetic strategies for circuit manipulation. A typical day consisted of morning lectures followed by hands-on laboratory practical sessions in the afternoon and evening with guest lecturers available to give one-on-one practical advice. The laboratory component of the course introduced students to state-of-the-art electrophysiological approaches for the study of

ion channels in their native environments. The course provided students with hands-on experience in using patch-clamp electrophysiology to examine single-channel activity in cultured cells, ion channel biophysics in acutely dissociated neurons, and synaptic integration, plasticity, and circuit dynamics in in vitro slice and in in vivo preparations. Different recording configurations were used (e.g., cell-attached, whole-cell dendritic and somatic patch, voltage- and current-clamp configurations) and the advantages and limitations of each method were discussed in relation to specific scientific questions. The course also provided practical experience in cellular and circuit manipulation techniques (i.e., pharmacological, electrophysiological, and optogenetic) both in vitro and in vivo.

This course was supported with funds provided by the Helmsley Charitable Trust and Howard Hughes Medical Institute.

PARTICIPANTS

Buelow, P., B.S., Emory University, Atlanta, Georgia
Fenk, L., Ph.D., Max-Planck Institute for Brain Research, Frankfurt/Main, Germany
Fisher, E., Ph.D., New York University Langone Medical Center, New York
Huang, M., B.S., The Scripps Research Institute, La Jolla, California
Lapato, A., Ph.D., University of California School of Medicine, Riverside
MacDonald, D.I., B.A., University College London, United Kingdom
Minder, J., M.S., New York University Sackler Institute, New York
Nagpal, L., Ph.D., Johns Hopkins University School of Medicine, Baltimore, Maryland
Phillips, H., B.S., SUNY Upstate Medical University, Syracuse, New York
Riad, M., B.S., The Rockefeller University, New York
Varma, A., B.Sc., National Centre for Biological Sciences, Bangalore, India
William, M., Ph.D., Dartmouth College, Lebanon, New Hampshire

SEMINARS

Carter, A., New York University, New York: Synaptic organization of the prefrontal cortex.
Cohen, J., The Johns Hopkins University, Baltimore, Maryland: Principles of extracellular electrophysiology.
Diamond, J., National Institutes of Health, Bethesda, Maryland: Synaptic integration underlies visual signaling in the retina.
Gasparini, S., Louisiana State University, New Orleans: Frequency-dependent input processing in hippocampal CA1 pyramidal.
Hull, C., Duke University, Durham, North Carolina: Cellular and circuit mechanisms of cerebellar synaptic processing.
Kammermeier, P., University of Rochester Medical Center, New York: Using channel modulation as a biosensor to examine mGluR function in neurons.
Lampert, A., RWTH Aachen University, Germany: Sodium channel mutations: Function, pharmacology, and disease modeling.
Larkum, M., Humboldt University of Berlin, Switzerland: Active cortical dendrites modulate perception.
Larsson, P., University of Miami, Florida: KCNE β subunits modulate voltage sensor movements of KCNQ1 channels.
Margrie, T., University College of London, United Kingdom: Whole-cell recording in vivo.
Nagel, G., Würzburg University, Germany: Channelrhodopsin et al.: Optogenetic actuators.
Nimigean, C., Weill Cornell Medical College, New York, New York: Potassium channels.
Plested, A., FMP Berlin, Germany: Single-channel recording.
Sjostrom, J., McGill University, Montreal, Quebec, Canada; Differential regulation of evoked and spontaneous release by presynaptic.
Spruston, N., Janelia Research Campus, Ashburn, Virginia: A crash course in neuronal excitability: How it works and why it matters.
Xu-Friedman, M., SUNY Buffalo, New York: Synaptic and intrinsic adaptations for enhancing reliability in an auditory.

Workshop on Pancreatic Cancer

June 6–11

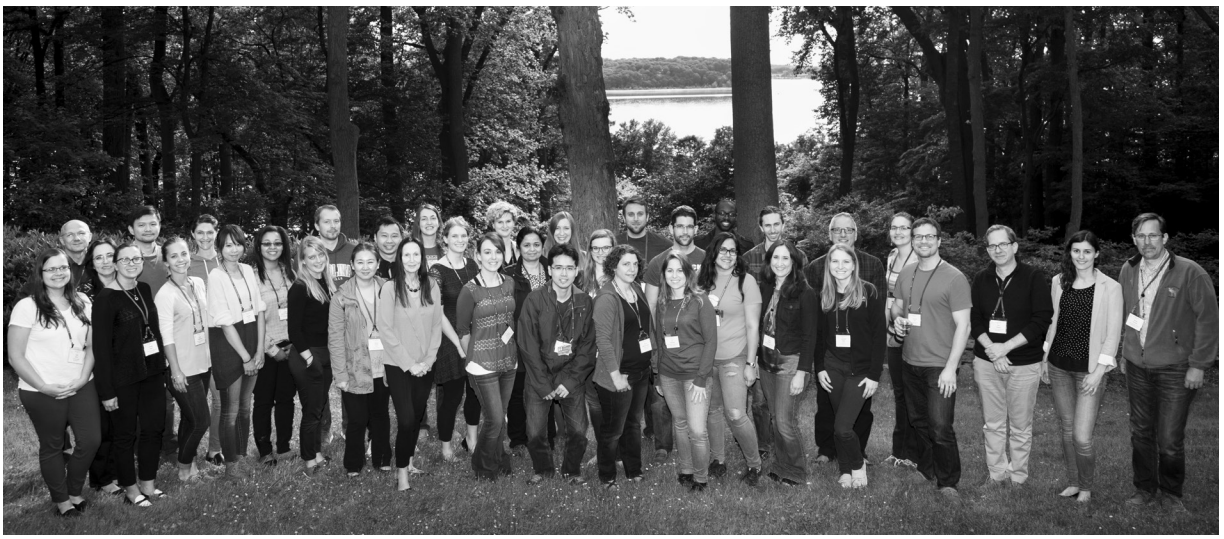
INSTRUCTORS
D. Bar-Sagi, New York University Langone Medical Center, New York
H. Crawford, University of Michigan, Ann Arbor
T. Hollingsworth, University of Nebraska, Omaha
D. Tuveson, Cold Spring Harbor Laboratory

ASSISTANTS
C. Chio, Cold Spring Harbor Laboratory
C-I. Hwang, Cold Spring Harbor Laboratory
T. Oni, Cold Spring Harbor Laboratory

Pancreatic cancer is one of the deadliest cancers. Tumors are often diagnosed at advanced stages of the disease and metastasize rapidly. This one-week discussion course provided a comprehensive overview of clinical and biological aspects of pancreatic cancer with special emphasis on disease diagnosis and management, molecular pathways involved in tumor development and progression, mechanism-based therapeutic strategies, advanced research tools, and ethical concerns. Attendees were able to interact with senior investigators on a one-to-one basis in an informal environment.

Topics included:

- Organ overview: Anatomy, physiology
- Clinical aspects of pancreatic cancer: Diagnosis, treatment
- Molecular genetics of pancreatic cancer: Gene signatures, predisposition syndromes
- Pathobiology of pancreatic cancer: Pathways, cell of origin, tumor microenvironment
- Tools and techniques: Mouse models, imaging, genomics, proteomics, metabolomics, bioinformatics
- Therapeutics: Target identification and validation, preclinical studies, clinical trial design
- Resources: Biobanks, funding strategies



This course was supported with funds provided by the Lustgarten Foundation.

PARTICIPANTS

Antal, C., Ph.D., Salk Institute for Biological Studies, La Jolla, California
Bojmar, L., Ph.D., Weill Cornell Medicine, New York, New York
Bray, J., B.S., University of Florida, Gainesville
Crawford, A., B.S., University of Nebraska Medical Center, Omaha
Dann, A., M.D., University of California, Los Angeles
Earley, K., Ph.D., Oregon Health and Science University, Portland
Fellmann, C., Ph.D., University of California, Berkeley
Gordillo, M., Ph.D., Weill Cornell Medical College, New York, New York
Ho, W., B.S., Massachusetts General Hospital, Boston
Hoffman, M., Ph.D., University of Michigan, Ann Arbor
Imasuen-Williams, I., B.A., Indiana University School of Medicine, Indianapolis
Leonhardt, L., M.Sc., University of California, San Francisco
Liang, G., Ph.D., Salk Institute for Biological Studies, La Jolla
Lytle, N., Ph.D., University of California, San Diego, La Jolla
Michaelis, K., B.A., Oregon Health & Science University, Portland
Nelson, B., B.S., University of Michigan, Ann Arbor
Papke, B., Ph.D., University of North Carolina, Chapel Hill
Pickup, M., Ph.D., University of California, San Francisco
Purohit, V., Ph.D., New York University, New York
Reindl, K., Ph.D., North Dakota State University, Fargo
Root, A., Ph.D., Memorial Sloan Kettering Cancer Center, New York, New York
Shaashua, L., Ph.D., The Weizmann Institute of Science, Rehovot, Israel
Vance, K., B.S., University of Nebraska Medical Center, Omaha
Velez, A., B.S., University of Michigan, Ann Arbor
Vonderhaar, E., B.S., Medical College of Wisconsin, Milwaukee
Zhang, Y., Ph.D., University of Oklahoma Health Science Center, Oklahoma City

SEMINARS

Simeone, D., New York University Langone Medical Center, New York: Surgery for pancreatic cancer.
O'Broin-Lennon, A.M., Johns Hopkins University School of Medicine, Baltimore, Maryland: Gastroenterology: Diagnosis and management of patients.
Al-Hawary, M., University of Michigan, Ann Arbor: Radiology: Imaging pancreatic neoplasms.
Maitra, A., University of Texas MD Anderson Cancer Center, Houston: Histological and molecular precursor neoplasms, pancreatitis, PDA, genetic progression series.
Herman, J., University of Texas MD Anderson Cancer Center, Houston: Radiation oncology for pancreatic cancer.
Ocean, A., Weill Cornell Medicine, New York, New York: Medical oncology approach to pancreatic cancer patients.
Tuveson, D., Cold Spring Harbor Laboratory: Mouse and organoid PDA models.
Bar-Sagi, D., New York University Langone Medical Center, New York: Ras and PDA.
Crawford, H., University of Michigan, Ann Arbor: The origin of pancreatic cancer.
Pasca di Magliano, M., University of Michigan, Ann Arbor: Inflammation in PDA evolution.
Xie, K., University of Texas MD Anderson Cancer Center, Houston: Transcriptional repressors in PDA origin.
Bailey, J., University of Texas Health Science Center at Houston, Houston: The ductal cell-PDA evolution controversy.
Saur, D., Technical University Münich, Department of Medicine II, Münich, Bavaria, Germany: Modeling and targeting pancreatic cancer and its microenvironment.
Holderfield, M., Frederick National Laboratory for Cancer Research, Frederick, Maryland: Targeting KRAS: Developing treatments for an undruggable protein.
Lewis, J., Memorial Sloan Kettering Cancer Center, New York, New York: Developing new imaging methods for PDA.
Egeblad, M., Cold Spring Harbor Laboratory: Using IVM to study pancreatic cancer.
Bardeesy, N., Massachusetts General Hospital Cancer Center, Boston: Epigenetic changes and pancreatic cancer.
Fearon, D.T., Cold Spring Harbor Laboratory: Cachexia and metabolism.
Vander Heiden, M., Massachusetts Institute of Technology, Cambridge: Metabolic changes in pancreatic cancer.
Wolpin, B., Dana-Farber Cancer Institute, Boston, Massachusetts: Personalized medicine for pancreatic cancer patients.
Miller, G., New York University Langone Medical Center, New York: Toll receptor inflammation in PDA.
Weaver, V., University of California, San Francisco: Tumor genotype, stromal phenotype, tissue tension, and PDAC aggression.
Sherman, M., Oregon Health & Science University, Portland: Metabolic and gene-regulatory functions of the pancreatic tumor microenvironment.

Hingorani, S., Fred Hutchinson Cancer Research Center, Seattle, Washington: Targeting the stroma in PDA.
Petersen, G., Mayo Clinic, Rochester, Minnesota: Screening high-risk kindreds for early PDA.
Chari, S., Mayo Clinic, Rochester, Minnesota: Late-onset diabetes and PDA.
Jaffee, E., Johns Hopkins University, Baltimore, Maryland: Vaccine-based combinatorial immunotherapy converts pancreatic cancer into an immunologic disease.
Hollingsworth, M.A. (Tony), University of Nebraska, Omaha: Early detection of pancreatic cancer.
Hanash, S.M., University of Texas MD Anderson Cancer Center, Houston: Search for PDAC markers from mouse to humans.
Alberto Diaz, L., Memorial Sloan Kettering Cancer Center, New York, New York: Circulating biomarkers: Cell-free DNA.
Stanger, B., University of Pennsylvania, Philadelphia: Modeling tumor heterogeneity in pancreatic cancer.
Kimmelman, A., New York University School of Medicine, New York: New drug targets in PDA metabolism.
Olive, K., Columbia University, New York: Targeting master regulators of pancreatic cancer malignancy.
Gallinger, S., Toronto General Hospital, Ontario, Canada: Molecular pathology of pancreatic cancer.
Beatty, G., University of Pennsylvania/Perelman School of Medicine, Philadelphia: Immunology of pancreatic cancer.

Mouse Development, Stem Cells, and Cancer

June 7–26

INSTRUCTORS **B. Allen,** University of Michigan Medical School, Ann Arbor
A. Ralston, Michigan State University, East Lansing

CO-INSTRUCTORS **T. Caspary,** Emory University School of Medicine, Atlanta, Georgia
C. Forsberg, University of California, Santa Cruz

ASSISTANTS **M. Echevarria Andino,** University of Michigan, Ann Arbor
T. Frum, Michigan State University, East Lansing
A. Lokken, Michigan State University, East Lansing
M. Scales, University of Michigan Medical School, Ann Arbor
S. Smith-Berdan, University of California, San Francisco
A. Snouffer, University of Georgia, Athens

This intensive lecture and laboratory course was designed for scientists interested in using mouse models to study mammalian development, stem cells, and cancer. The lecture portion of the course, taught by leaders in the field, provided the conceptual basis for contemporary research in embryogenesis; organogenesis in development and disease; embryonic, adult, and induced pluripotent stem cells; and cancer biology.

The laboratory and workshop portions of the course provided hands-on introduction to engineering of mouse models, stem cell technologies, and tissue analyses. Experimental techniques included genome editing by CRISPR-Cas9; pronuclear microinjection; isolation and culture/manipulation of pre- and postimplantation embryos; embryo transfer; embryo electroporation and

roller bottle culture; chimera generation; generation and differentiation of mouse embryonic stem cells and fibroblasts; vibratome and cryosectioning; in situ RNA hybridization; immunostaining; FACS sorting and analysis of hematopoietic stem cells; skeletal preparation; organ explant culture; and fluorescent imaging, including live time-lapse microscopy.

This course was supported with funds provided by the National Cancer Institute.

PARTICIPANTS

Bromfield, E., Ph.D., The University of Newcastle, Callaghan, New South Wales, Australia
Brown, D., Ph.D., Duke University Medical Center, Durham, North Carolina
Camargo Ortega, G., Ph.D., Helmholtz Zentrum Muenchen, Neuherberg/Münich, Germany
Gigante, E., M.S., Emory University, Atlanta, Georgia
Goodwin, K., M.Sc., Princeton University, New Jersey
Lea, R., M.Bioch., The Francis Crick Institute, London, United Kingdom
Lo Cascio, C., B.Sc., Barrow Neurological Institute, Phoenix, Arizona
Müller, I., M.S., University of Copenhagen, Denmark
Ngo, J., B.A., Case Western Reserve University, Cleveland, Ohio
Prieto, L., B.S., Mayo Clinic, Rochester, Minnesota
Ramakrishnan, A., B.Tech., University of Michigan, Ann Arbor
Song, C., Ph.D., University of Massachusetts Medical School, Worcester
Sutherland, J., Ph.D., The University of Newcastle, Callaghan, Australia
Yuan, S., Ph.D., Yale University School of Medicine, New Haven, Connecticut

SEMINARS

Allen, B., University of Michigan Medical School, Ann Arbor: Hedgehog signaling in embryonic development, adult tissue homeostasis, and cancer.
Blelloch, R., University of California Medical Center, San Francisco: Loss and gain of pluripotency.
Brugmann, S., Cincinnati Children's Hospital, Ohio: Primary cilia in development and disease.
Caspary, T., Emory University School of Medicine, Atlanta, Georgia: Listening to the mouse mutants: Neural development.
Clark, A., University of California, Los Angeles: Methylation reprogramming in primordial germ cells.
Costantini, F., Columbia University, Pelham, New York: Branching morphogenesis during kidney development.
Forsberg, C., University of California, Santa Cruz: Hematopoietic stem cells.
Lacy, E., Memorial Sloan Kettering Cancer Center, New York, New York: Gastrulation and early organogenesis.
Laird, D., University of California Medical Center, San Francisco: Lineage and fate in the germline.

Mouse anniversary group

Lewandoski, M., National Cancer Institute/NIH, Frederick, Maryland: Mouse genetics technologies.
Mager, J., University of Massachusetts, Amherst: Regulation of genomic imprinting during early development.
Magnuson, T., University of North Carolina, Chapel Hill: Genome-wide dynamics of chromatin modifiers.
Ralston, A., Michigan State University, East Lansing: Stem cells and the first cell fate decisions in mouse development.
Robertson, E., University of Oxford, United Kingdom: Axis formation.
Trainor, P., Stowers Institute for Medical Research, Kansas City, Missouri: Neural crest cells: Their roles in evolution, development, disease, and cancer.
Tuveson, D., Cold Spring Harbor Laboratory: Pancreatic cancer models.
Witte, O., University of California, Los Angeles: Finding targets for cancer therapy.
Witkowski, J., Cold Spring Harbor Laboratory: Ethics, reproducibility, and vigor.

Metabolomics

June 10–26

INSTRUCTORS **A. Caudy,** University of Toronto, Ontario, Canada
J. Cross, Memorial Sloan Kettering Cancer Center, New York, New York
E. Gottlieb, Cancer Research UK Beatson Institute, Glasgow, United Kingdom
A. Rosebrock, Stony Brook University, New York

ASSISTANTS **I. Abramovich,** Technion Institute of Technology, Haifa, Israel
M. Berisa, Memorial Sloan Kettering Institute, New York, New York
J. Hanchard, University of Toronto, Ontario, Canada
G. MacKay, Cancer Research UK Beatson Institute, Glasgow, United Kingdom
D. Sumpton, Cancer Research UK Beatson Institute, Glasgow, United Kingdom

Metabolomics involves the systematic interrogation of the abundance of small chemical molecules (metabolites) within cells, tissues, organs, and organisms. In parallel with high-throughput technologies that facilitate genomic, transcriptomic, and proteomic analyses of cellular and organismal physiology, technologies for metabolite profiling represent an important source of information about the dynamic state of the cell or tissue that is relevant in both health and disease.

LC-MS metabolomics was the primary focus of this course and was applied for both targeted and untargeted analyses of endogenous metabolites and in vitro enzyme reactions. We used approaches for steady-state measurement of metabolite levels as well as assessment of metabolite flux. To complement these LC-MS analyses, students performed experiments involving other methodologies. Shorter-term activities with other methodologies included GC-MS, polarimetric and Seahorse measurement of oxygen consumption, FRET sensors and/or MitoTracker measurements, enzymatic techniques for metabolite measurement, and uptake experiments.

The consistent and extended application of LC-MS reflected the emphasis of the course, and the exposure to other methodologies allowed students to appreciate the utility and complementarity of these methods.

Objectives for Students:

- Quantitative and qualitative analyses of LC-MS data using currently available tools (vendor software, Rosebrock tools, XCMS online, Agilent Profinder/Genespring).
- Understand common interferences and limitations of LC-MS and GC-MS analysis.
- Recognize key issues in experimental design and sample preparation for metabolomics.
- Awareness of major biochemical pathways active in commonly used cell types.
- Familiarity with methods for determining different types of oxygen consumption.

Proposed Lab Exercises:

- Full-scan experiment on knockout/drug treatment in which students identified significantly changed metabolites and used MS/MS fragmentation and other methods for identification.
- Experimental treatments were selected in which the discovery of a phenotype was possible with only one of several analytic/separation methodologies.
- Identification of a significantly changed metabolite in blood/plasma/urine and development of a targeted method for its analysis by QQQ, including determination of LOD/LOQ/linearity and other appropriate method validation.
- Measurement of metabolite flux by pulse labeling (i.e., kinetic flux profiling).
- Enzyme assay to determine V_{max} and K_m (measure on MSD?).
- Enzymatic synthesis of a compound (e.g., sedoheptulose bisphosphate, or ribose-1-phosphate) and purification by mass-based or HPLC fractionation.
- Non-LC-MS methodologies: Roche kits for lactate and glucose measurement in culture supernatant; seahorse experiment, including all drugs for uncoupled, etc.; FRET experiment (by cytometry and/or microscopy) for NADH levels and proton gradient–sensitive Mitotracker staining; nutrient uptake by 14C (potentially amino acids?); and glycolytic flux in mammalian cells by 3H.

Students received hands-on training on Agilent QTOF and Thermo Orbitrap, Q-exactive and Vantage triple quadrupole, and Waters SYNAPT G2-S mass spectrometers.

This course was supported with funds provided by the National Institute of General Medical Sciences, Helmsley Charitable Trust, and Howard Hughes Medical Institute.

PARTICIPANTS

Bonem, M., B.S., University of Texas, Austin
Chan, S.H., M.S., Whitehead Institute for Biomedical Research, Cambridge, Massachusetts
Cox, L., Ph.D., Brigham & Women's Hospital, Harvard Medical School, Boston, Massachusetts
Dall, K., M.S., University of Southern Denmark, Odense
Hancock, S., Ph.D., University of Wollongong, New South Wales, Australia
Imholz, N., M.Sc., Delft University of Technology, Netherlands
Jackson, S., Ph.D., National Institute of Standards and Technology, Gaithersburg, Maryland
Kalia, V., M.P.H., Emory University, Rollins School of Public Health, Atlanta, Georgia
Kumar, A., B.S., University of California, San Diego, La Jolla
Marin-Valencia, I., M.D., The Rockefeller University, New York, New York
Onwuha-Ekpete, L., B.Sc., Florida Atlantic University, Boca Raton
Parker, A., Ph.D., National Cancer Institute, Bethesda, Maryland
Shaw, T., Ph.D., St Jude Children's Research Hospital, Memphis, Tennessee
Singh, R., Ph.D., Florida State University, Tallahassee
Stepanova, V., M.S., Skolkovo Institute of Science and Technology, Moscow, Russia
Stor, D., B.A., The Rockefeller University, New York, New York

SEMINARS

Amador-Noguez, D., University of Wisconsin, Madison: Limited thermodynamic driving force in glycolysis of cellulolytic clostridia.

Clasquin, M., Pfizer, Boston, Massachusetts: Mass spectrometry–based metabolomics to gain mechanistic insights around drug targets.

DeBerardinis, R., University of Texas Southwestern Children's Medical Center, Dallas: Assessing metabolic flux in live tumors in mice and humans.

Fan, W-M.T., University of Kentucky, Lexington: NMR profiling of metabolites and their 13C isotopomers in stable isotope-resolved metabolomics studies.

Gross, S., Weill Cornell Medical College, New York, New York: Untargeted metabolite profiling to discover mechanisms of drug actions and gene functions and to screen for inborn errors of metabolism.

Keshari, K., Memorial Sloan Kettering Cancer Center, New York, New York: Hyperpolarized magnetic resonance to investigate metabolism.

Metallo, C., University of California, San Diego, La Jolla: Exploring links between amino acid and lipid metabolism in cancer and diabetes.

Patti, G., Washington University, St. Louis, Missouri: Promoting anabolic metabolism in cancer with altered redox balance.

Rhee, K., Weill Cornell Medical College, New York, New York: Mycobacterial metabolomics: Chemical biology at the intersection of pathogen biology and drug development.

Vander Heiden, M., Massachusetts Institute of Technology, Cambridge: Determinants of metabolic dependencies in tumors.

Witkowski, J., Cold Spring Harbor Laboratory: Ethics, reproducibility, and vigor.

Vision: Linking Circuits, Perception, and Behavior

June 15–29

INSTRUCTORS **F. Briggs,** University of Rochester Medical Center, New York
A. Huberman, Stanford University School of Medicine, Palo Alto, California

CO-INSTRUCTOR **O. Dhande,** Stanford University School of Medicine, Palo Alto, California

ASSISTANT **L. Salay,** Stanford University School of Medicine, California

The purpose of this course was to bring together students and faculty for in-depth and high-level discussions of modern approaches for probing how specific cell types and circuits give rise to defined categories of perception and action. It was also designed to address novel strategies aimed at overcoming diseases that compromise sensory function.

The visual system is the most widely studied sensory modality. Recently, three major shifts have occurred in the field of neuroscience. First, because of the large array of genetic techniques available in mice and the relative ease of imaging and recording from the cortex of small rodents, the mouse visual system has become a premiere venue for attacking the fundamental unresolved question of how specific cells and circuits relate to visual performance at the receptive field and whole-animal level. Second, genetic and viral methods have evolved to the point where neurophysiologists can directly probe the role of defined circuits in species such as macaque monkeys, thus bridging the mechanism-cognition gap. Third, the field of visual neuroscience is rapidly paving the way for widespread clinical application of stem cell, gene therapy, and prosthetic devices to restore sensory function in humans.

The time is ripe to build on the classic paradigms and discoveries of visual system structure, function, and disease to achieve a deep, mechanistic understanding of how receptive fields are organized and filter sensory information, how that information is handled at progressively higher levels of neural processing, and how different circuits can induce defined categories of percepts and behaviors in the healthy and diseased brain.

This course was supported with funds provided by the National Eye Institute, Helmsley Charitable Trust, and Howard Hughes Medical Institute.

PARTICIPANTS

Akrouh, A., Ph.D., Columbia University, New York, New York
Archer, D., Ph.D., University of California, Davis
Bard, A., B.Sc., King's College London, United Kingdom
Care, R., B.S., University of California, San Francisco
Cazemier, L., M.Sc., Netherlands Institute for Neuroscience, Amsterdam, Netherlands
Chaney, S., Ph.D., University of California, San Francisco
Cheung, V., B.S., University of California, San Francisco
DePiero, V., B.S., University of Louisville School of Medicine, Louisville, Kentucky
Dougherty, K., B.S., Vanderbilt University, Nashville, Tennessee
Jung, H., Ph.D., Stanford University, California
Kirchgessner, M., B.A., Salk Institute for Biological Studies, La Jolla, California
Liu, L. (Dave), Ph.D., McGill University, Montreal, Quebec, Canada
Mazade, R., Ph.D., State University of New York College of Optometry, New York
Parker, P., Ph.D., University of Oregon, Eugene
Pearson, C., B.S., University of Cambridge, United Kingdom
Prigge, C., Ph.D., Duke University, Durham, North Carolina
Reinhard, K., Ph.D., Neuro-Electronics Research Flanders, Leuven, Belgium
Richards, S., B.S., Brandeis University, Waltham, Massachusetts
Salmon, A., B.S., Medical College of Wisconsin, Milwaukee
Savier, E., Ph.D., Northwestern University, Evanston, Illinois
Sengupta, A., B.S., National Eye Institute/NIH Grad Partnerships Program (University of Paris-6), Bethesda, Maryland
Shah, S., B.S., Stanford University, Palo Alto, California
Srivastava, M., M.Sc., University of Debrecen, Hungary
Wailes-Newson, K., B.Sc., University of York, United Kingdom
Yoo, M., M.S., The University of Chicago, Illinois
Zarei Eskikand, P., M.S., The University of Melbourne, Parkville, Australia

SEMINARS

Agarwal, N., National Eye Institute, Rockville, Maryland: The NEI audacious goals initiative (given with S. Becker and T. Greenwell). NIH/NEI training fellowship opportunities for postdocs and graduate students (given with S. Becker and T. Greenwell). Alternative career paths for biomedical students (given with S. Becker and T. Greenwell).
Becker, S., National Institutes of Health, Bethesda, Maryland: The NEI audacious goals initiative (given with N. Agarwal and T. Greenwell). NIH/NEI training fellowship opportunities for postdocs and graduate students (given with N. Agarwal and T. Greenwell). Alternative career paths for biomedical students (given with N. Agarwal and T. Greenwell).
Briggs, F., Geisel School of Medicine at Dartmouth, Lebanon, New Hampshire: Visual system, overview (given with A. Huberman). Career development Part 1 (given with A. Huberman and O. Dhande). Career development Part 2 (given with A. Huberman and O. Dhande). Career development Part 3 (given with A. Huberman and O. Dhande).
Carroll, J., Medical College of Wisconsin Eye Institute, Milwaukee: Color vision in health and disease.
Churchland, A., Cold Spring Harbor Laboratory: State-dependent influences on visual processing.
Conway, B., National Eye Institute, Bethesda, Maryland: Color.
Dhande, O., Stanford University School of Medicine, Palo Alto, California: Retina cell types. Career development Part 1 (given with A. Huberman and F. Briggs). Career development Part 2 (given with A. Huberman and F. Briggs). Career development Part 3 (given with A. Huberman and F. Briggs).
Dunn, F., University of California, San Francisco: Gain control in the mammalian retinal development and disassembly of the visual system's first synapse.
Fitzpatrick, D., Max Planck Florida Institute for Neuroscience, Jupiter: Visualizing cortical maps and their development with cellular and synaptic resolution.
Gallant, J., University of California, Berkeley: Natural vision.
Gottlieb, J., Columbia University, New York, New York: Eye movements and attention.
Greenwell, T., National Eye Institute, Rockville, Maryland: The NEI audacious goals initiative (given with N. Agarwal

and S. Becker). NIH/NEI training fellowship opportunities for postdocs and graduate students (given with N. Agarwal and S. Becker). Alternative career paths for biomedical students (given with N. Agarwal and S. Becker).

Hirsch, J., University of Southern California, Los Angeles: Receptive field transformations.

Hofer, S., University of Basel, Switzerland: Functional architecture and cortical specializations.

Horton, J., University of California, San Francisco: Visual system disease II.

Huberman, A., Stanford University School of Medicine, Palo Alto, California: Visual systems, overview (given with F. Briggs). Visual system disease I. Career development Part 1 (given with F. Briggs and O. Dhande). Career development Part 2 (given with F. Briggs and O. Dhande). Career development Part 3 (given with F. Briggs and O. Dhande).

Kiorpes, L., New York University, New York: Visual system disease II.

Krauzlis, R., National Eye Institute/NIH, Bethesda, Maryland: Subcortical control of eye movements and attention.

Mitchell, J., Systems Neurobiology Lab, The Salk Institute, La Jolla, California: Neural mechanisms of spatial attention and active vision in the marmoset.

Moore, T., Stanford University, California: Attention, eye movements, and goal-directed behavior.

Movshon, J.A., New York University, New York: Receptive field transformations.

Nielsen, K., Johns Hopkins University, Baltimore, Maryland: Visual system overview.

Salzman, C.D., Columbia University, New York, New York: Multisensory integration.

Sherman, S.M., The University of Chicago, Illinois: Subcortical control of eye movements and attention.

Usrey, W.M., University of California, Davis: Functional properties of neural circuits for vision.

Witkowski, J., Cold Spring Harbor Laboratory: Ethics seminar.

Single Cell Analysis

June 30–July 13

INSTRUCTORS **D. Chenoweth,** University of Pennsylvania, Merion Station
M. McConnell, University of Virginia School of Medicine, Charlottesville
G. Yeo, University of California, San Diego, La Jolla

CO-INSTRUCTORS **O. Botvinnik,** Chan Zuckerberg Biohub, Fremont, California

ASSISTANTS **C. (Jay) Aonbangkhen,** University of Pennsylvania, Philadelphia
A. Domissy, University of California, San Diego, La Jolla
M. Haakenson, University of Virginia School of Medicine, Charlottesville
E. Wheeler, University of California San Diego, La Jolla
M. Wolpert, University of Virginia, Charlottesville
D. Wu, University of Pennsylvania, Philadelphia

The goal of this two-week course was to familiarize students with cutting-edge technologies for characterization of single cells. Modules of the course were taught by scientists with expertise in distinct areas of single-cell analysis. Topics covered included quantitative single-cell analysis by RNA-Seq, genomic DNA analysis, proteomics, and metabolomics. Multiple nucleic amplification methodologies including droplet-based RNA-Seq, MALBAC, and MDA were used. In addition, students were instructed in basic bioinformatic analysis of next-generation sequencing data.

Below are listed the topics included in this course:

- Single-cell genome, transcriptome, and proteome measurement.
- Introductory next-generation sequencing data analysis.
- Photoactivatable single-cell probes.
- Single-cell mass spectrometry/soft X-ray tomography.

This course was supported with funds provided by the National Institute of General Medical Sciences, Howard Hughes Medical Institute, and Helmsley Charitable Trust.

PARTICIPANTS

Aquino Nunez, W., B.S., Kennesaw State University, Georgia
Blosser, T., Ph.D., Broad Institute, Cambridge, Massachusetts
Brandebura, A., B.S., West Virginia University, Morgantown
Chen, N., Ph.D., Massachusetts Institute of Technology, Cambridge
Deslauriers, A., M.Sc., Memorial Sloan Kettering Cancer Center, New York, New York
Diaz, G., M.S., Stanford University, California
Grunwald, S., Ph.D., Charité Medical Faculty, Berlin, Germany
Hou, Y., B.S., Washington University, St. Louis, Missouri
Kashyap, A., M.S., ETH Zürich, Basel, Switzerland
Kim, T.H., Ph.D., Case Western Reserve University, Cleveland, Ohio
Papke, B., Ph.D., University of North Carolina, Chapel Hill
Sharma, A., Ph.D., Mayo Clinic, Rochester, Minnesota
Simon, R., B.S., University of North Carolina, Chapel Hill
Torkenczy, K., B.S., Oregon Health and Science University, Portland
Tsai, J., M.D./Ph.D., Boston Children's Hospital, Boston, Massachusetts
Yang, S., Ph.D., The Jackson Laboratory, Bar Harbor, Maine
Yu, K., Ph.D., St. Michael's Hospital, Toronto, Ontario, Canada
Zhang, H., Ph.D., Columbia University Medical Center, New York, New York

SEMINARS

Allbritton, N., University of North Carolina, Chapel Hill: Microdevices to assay single cells.
Bintu, L., Stanford University, California: Dynamics of epigenetic regulation using time-lapse microscopy.
Chenoweth, D., University of Pennsylvania, Merion Station: New chemical tools for cell biology.
Eberwine, J., University of Pennsylvania, Philadelphia: Multimodal single-cell ′omics: Are all cells unicorns?
Herr, A., University of California, Berkeley: High-selectivity single-cell protein assays using electrophoretic cytometry.
Larabell, C., University of California, San Francisco: CT scans of single cells.
McConnell, M., University of Virginia School of Medicine, Charlottesville: Brain somatic mosaicism.
Rubakhin, S., University of Illinois, Urbana: Introduction to mass spectrometry.
Sandberg, R., Karolinska Institutet, Stockholm, Sweden: Single-cell RNA sequencing.
Shultz, C., Oregon Health & Science University, Portland: Imaging tools for studying signaling networks.
Sims, P., Columbia University, New York, New York: Applications and analysis of large-scale, single-cell RNA-Seq (joint with CSHL's Statistical Methods for Functional Genomics Course).
Trotta, N., Cell Microsystems, Inc., Research Triangle Park, North Carolina: The CellRaft AIR™ system: Advanced and applications.
Yeo, G., University of California, San Diego, La Jolla: Single-cell alternative splicing analysis.
Zhuang, X., Harvard University, Cambridge, Massachusetts: Single-cell transcriptome and chromosome imaging.

Statistical Methods for Functional Genomics

June 30–July 13

INSTRUCTORS **H. Bussemaker,** Columbia University, New York, New York
S. Davis, National Institutes of Health, Columbia, Maryland
R. Irizarry, Dana-Farber Cancer Institute, Boston, Massachusetts
T. Lappalainen, New York Genome Center/Columbia University, New York

ASSISTANTS **M. Brandt,** Columbia University, New York, New York
S. Castel, Columbia University, New York, New York
V. FitzPatrick, Columbia University, New York, New York
P. Kimes, Dana-Farber Cancer Institute, Boston, Massachusetts
K. Korthauer, Dana-Farber Cancer Institute, Boston, Massachusetts
J. Kribelbauer, Columbia University, New York, New York
H.T. Rube, Columbia University, New York, New York

Over the past decade, high-throughput assays have become pervasive in biological research due to both rapid technological advances and decreases in overall cost. To properly analyze the large data sets generated by such assays and thus make meaningful biological inferences, both experimental and computational biologists must understand the fundamental statistical principles underlying analysis methods. This course was designed to build competence in statistical methods for analyzing high-throughput data in genomics and molecular biology.

Below are listed the topics included in the course:

- The R environment for statistical computing and graphics.
- Introduction to Bioconductor.
- Review of basic statistical theory and hypothesis testing.
- Experimental design, quality control, and normalization.
- High-throughput sequencing technologies.
- Expression profiling using RNA-Seq and microarrays.
- In vivo protein binding using ChIP-seq.

- High-resolution chromatin footprinting using DNase-seq.
- DNA methylation profiling analysis.
- Integrative analysis of data from parallel assays.
- Representations of DNA binding specificity and motif discovery algorithms.
- Predictive modeling of gene regulatory networks using machine learning.
- Analysis of posttranscriptional regulation, RNA binding proteins, and microRNAs.

Detailed lectures and presentations by instructors and guest speakers were combined with hands-on computer tutorials. The methods covered in the lectures were applied to example high-throughput data sets.

This course was supported with funds provided by the National Institute of General Medical Sciences.

PARTICIPANTS

Barrows, D., Ph.D., The Rockefeller University, New York, New York
Bowles, K., Ph.D., Icahn School of Medicine at Mount Sinai, New York
Brown, C., Ph.D., Memorial Sloan Kettering Cancer Center, New York, New York
Carmona, L., Ph.D., Columbia University, New York, New York
Celen, I., B.S., University of Delaware, Newark
Colaneri, A., Ph.D., University of North Carolina, Chapel Hill
Craig, J., B.A., Virginia Commonwealth University, Richmond
Fischer, J., Ph.D., Francis Crick Institute/University College London, United Kingdom
Gedman, G., B.A., The Rockefeller University, New York, New York
Getzler, A., B.A., The Scripps Research Institute-Scripps Florida, Jupiter
Giusti-Rodríguez, P., Ph.D., University of North Carolina, Chapel Hill
Hossain, M., Ph.D., University of Florida, Gainesville
Ietswaart, R., Ph.D., Harvard Medical School, Boston, Massachusetts
Koguchi, Y., Ph.D., Earle A. Chiles Research Institute, Portland, Oregon
Lai, C-Y., Ph.D., Salk institute for Biological Studies, La Jolla, California
Leiendecker, L., M.S., Research Institute of Molecular Pathology, Vienna, Austria
Lesseur, C., Ph.D., Icahn School of Medicine at Mount Sinai, New York
Matusiak, M., Ph.D., Stanford University, Stanford, California
Maurer, C., M.D., Columbia University, New York, New York
Negi, S., B.Tech., University of Illinois, Urbana-Champaign
Sanchez-Rivera, F., Ph.D., Memorial Sloan Kettering Cancer Center, New York, New York
Stoyanova, E., B.S., The Rockefeller University, New York, New York
Tsukamoto Kagohara, L., Ph.D., Johns Hopkins University, Baltimore, Maryland

SEMINARS

Bussemaker, H., Columbia University, New York, New York: Normal distribution and multiple testing. Scoring differential expression. Gene ontology scoring. Accurate quantification of DNA recognition and methylation readout by transcription factors. Basic motif discovery. Intro to ChIP-seq analysis. Weight matrices. Linear models of regulatory networks.
Churchman, S., Harvard Medical School, Boston, Massachusetts: Across the genome and across the cell, gene expression regulation at high resolution.
Davis, S., National Institutes of Health, Columbia, Maryland: Introduction to R. Bioconductor overview. Basics of RNA-Seq analysis. Scalable, cloud-based computation for genomics. Cloud genomics workshop.
Irizarry, R., Dana-Farber Cancer Institute, Boston, Massachusetts: Exploratory data analysis and data visualization. Systematic errors and batch effect in high-throughput biology. Introduction to high-dimensional data. Principle component analysis and its applications. Missing data and technical variability in single-cell RNA-sequencing experiments.
Kundaje, A., Stanford University, California: Molecular dialogues between pollen and Pistil. Learning the regulatory architecture of the human genome.
Lappalainen, T., New York Genome Center/Columbia University, New York: Introduction to NGS. Introduction to eQTLs. Allele-specific expression. Functional variation in the human genome: Lessons from the transcriptome.
Love, M., University of North Carolina, Chapel Hill: RNA-Seq differential expression. RNA-Seq differential expression and batch effects.
Rube, T., Columbia University, New York, New York: Basics of linear regression.
Sims, P., Columbia University, New York, New York: Applications and analysis of large-scale, single-cell RNA-Seq (joint with Single-Cell Course).
Zhu, A., University of North Carolina, Chapel Hill: Rmarkdown and Git.

Advanced Techniques in Molecular Neuroscience

June 30–July 15

INSTRUCTORS
C. Lai, Indiana University, Bloomington
J. LoTurco, University of Connecticut, Storrs
A. Schaefer, Icahn School of Medicine at Mount Sinai, New York

ASSISTANTS
P. Ayata, Icahn School of Medicine at Mount Sinai, New York
A. Badimon, Icahn School of Medicine at Mount Sinai, New York
A. Battison, University of Connecticut, Willimantic
M.K. Duff, Mount Sinai School of Medicine, New York
J. Fang, University of Connecticut, Willington
E. Perez, Indiana University, Bloomington
M. Pozsgai, Indiana University, Bloomington
M. Sakai, Indiana University, Bloomington
J. Sullivan, Icahn School of Medicine at Mount Sinai, New York
N. Volk, Icahn School of Medicine at Mount Sinai, New York

This laboratory and lecture course was designed to provide neuroscientists at all levels with a conceptual and practical understanding of several of the most advanced techniques in molecular neuroscience. The course curriculum was divided into three sections: an extensive and up-to-date set of laboratory exercises, daily lectures covering the theoretical and practical aspects of the various methods used in the laboratory, and a series of evening research seminars. The informal and interactive evening lectures were given by leading molecular neuroscientists and served to illustrate the ways in which the various experimental approaches have been used to advance specific areas of neurobiology. In this year's course, the laboratory portion included topics such as an introduction to the design and use of animal virus vectors in neurobiology; the use of CRISPR genome editing and RNAi approaches for regulating the expression of specific genes in neurons; practical exercises in gene delivery systems, including mammalian cell infection and transfection and

electroporation techniques for targeted gene transfer in vivo; an introduction to overall strategies, use, and design of BAC transgenic vectors; RT-PCR analyses; assays of chromatin and chromatin structure in neurons; and mRNA isolation from specified neural subtypes by TRAP.

This course was supported with funds provided by the National Institute of Mental Health.

PARTICIPANTS

Cantin, L., B.S., The Rockefeller University, New York, New York
Chan, U., B.S., Duke University, Durham, North Carolina
Jakobsen, E., M.S., University of Copenhagen, Denmark
del Carmen Krawczyk, M., Ph.D., University of Buenos Aires, Caba, Argentina
Krishnan, S., Ph.D., Universität Münster, Germany
Lee, K-S., B.S., Max Planck Florida Institute, Jupiter, Florida
Lim, H., Ph.D., CUNY, Hunter College and the Graduate Center, New York
Marshall, K., Ph.D., The Scripps Research Institute, La Jolla, California
Moreno Blas, D., B.S., National Autonomus University of Mexico, Mexico City
Moura, D., M.S., Ph.D., Universidade Federal do Rio Grande do Norte, Natal, Brazil
Reynolds, L., B.Sc., McGill University, Montreal, Quebec, Canada
Rupprecht, P., Diplom, Friedrich Miescher Institute, Basel, Switzerland
Suter, T., Ph.D., Brown University, Providence, Rhode Island
Toskas, K., M.Sc., Karolinska Institute, Stockholm, Sweden
Turner, T., Ph.D., University of Washington, Seattle
Zhang, Q., Ph.D., Riken, Wako, Japan

SEMINARS

Akbarian, S., Icahn School of Medicine at Mount Sinai, New York: Genome-scale 3D genome mappings in the nervous system.
Cai, D., University of Michigan, Ann Arbor: Multispectral labeling and reconstructing neural circuits at the single synapse.
Casaccia, P., Icahn School of Medicine at Mount Sinai, New York: The evolving story of oligodendrocyte lineage cells.
Cong, L., Broad Institute of MIT/Harvard, Cambridge, Massachusetts: CRISPR genome engineering module.
Darnell, R., The Rockefeller University/HHMI/NY Genome Center, New York: Applying advanced techniques in molecular neuroscience to the human brain.
Desmond, N., National Institute of Mental Health, Washington, D.C.: News from the NIH.
Eberwine, J., University of Pennsylvania, Philadelphia: Are all cells unicorns?: Multimodal analysis of in vivo and in vitro single cells.
Gradinaru, V., California Institute of Technology, Pasadena: Optogenetic, tissue clearing, and systemic viral vector approaches to understand and influence whole-animal physiology and behavior.
Haas, K., University of British Columbia, Vancouver, Canada: Single-cell electroporation for targeted in vivo transfection.
Harwell, C., Harvard Medical School, Boston, Massachusetts: Cellular and molecular mechanism of cortical circuit development.
Jaffrey, S., Weill Cornell Medical College, New York, New York: Epitranscriptomics: Mapping and function of dynamic nucleotide modifications in mRNA.
Kanadia, R., University of Connecticut, Storrs: Combined detection of RNA by ISH followed by proteins by IF and EdU by ClickIT.
Kenny, P., Icahn School of Medicine at Mount Sinai, New York: Mechanisms of nicotine craving.
Luikart, B., Geisel School of Medicine at Dartmouth, Lebanon, New Hampshire: Understanding the neurobiological basis of autism using Pten genetic models.
Maher, B., Lieber Institute for Brain Development/ JHMI, Baltimore, Maryland: Practical uses of in utero electroporation to study cortical development.
Schmidt, E., The Rockefeller University, New York, New York: Molecular phenotyping of distinct cortical cell types using the translating ribosome affinity purification (TRAP) approach.
Sehgal, A., University of Pennsylvania, Philadelphia: Regulation of sleep in *Drosophila* (joint with *Drosophila* Neurobiology course).
Silver, D., Duke University Medical Center, Durham, North Carolina: Techniques in slice culture and clonal analysis to investigate cortical development.
Witkowski, J., Cold Spring Harbor Laboratory: Ethics, reproducibility, and rigor.

Drosophila Neurobiology: Genes, Circuits, and Behavior

June 30–July 20

INSTRUCTORS
K. Kaun, Brown University, Providence, Rhode Island
A. Keene, Florida Atlantic University, Jupiter
C-H. Lee, National Institute of Child Health and Human Development/NIH, Bethesda, Maryland
S. Pulver, University of St. Andrews, United Kingdom

ASSISTANTS
J. Jonaitis, University of St. Andrews, United Kingdom
J. MacLeod, University of St. Andrews, United Kingdom
K. Nunez, Brown University, Providence, Rhode Island
E. Petruccelli, Brown University, Providence, Rhode Island
B-M. Song, National Institutes of Health, Bethesda, Maryland
B. Stahl, Florida Atlantic University, North Palm Beach
C-Y. Ting, National Institute of Child Health and Human Development/NIH, Bethesda, Maryland

This laboratory/lecture course was intended for researchers at all levels from beginning graduate students through established primary investigators who want to use *Drosophila* as an experimental system for nervous system investigation. This three-week course was designed to introduce students to a wide variety of topics and techniques, including the latest approaches for studying nervous system development, activity, and connectivity, as well as complex behaviors and disease models. Daily research seminars presented comprehensive overviews of specific subfields of nervous system development or function or focused on state-of-the-art techniques and approaches in *Drosophila* neuroscience. Expert guest lecturers discussed their findings and approaches and brought along their own assays and techniques for students to learn in the laboratory part of the course. The hands-on portion of the course was centered on inquiry-based projects, using the different morphological and physiological measurements and behavioral paradigms learned in the course. This included molecular-genetic analyses, immunocytochemistry, recording of activity

using electrophysiology and genetically encoded calcium indicators, optogenetic and thermogenetic control of neural activity, as well as numerous quantitative behavioral measures. Collectively, the course provided a comprehensive and practical introduction to modern experimental methods for studying the neural basis of behavior in *Drosophila*.

This course was supported with funds provided by the National Institute on Drug Abuse, National Science Foundation, Helmsley Charitable Trust, and Howard Hughes Medical Institute.

PARTICIPANTS

Bridi, J., M.Sc., King's College London, United Kingdom
de Andres Bragado, L., M.Sc., University of Fribourg, Switzerland
De Nobrega, A., B.S., Florida State University, Tallahassee
Goto, J., Ph.D., California State University, Fresno
Herbert, M., B.Sc., University of Oxford, United Kingdom
Lebhardt, F., Ph.D., New York University, New York
Leopold Cunningham, K., B.S., Massachusetts Institute of Technology, Cambridge
Schenk, J., B.S., University of Maryland, College Park
Snell, N., B.A., Brown University, Providence, Rhode Island
Tumkaya, T., M.Sc., National University of Singapore/AStar, Singapore
Vernon, S., B.Sc., University of Manchester, United Kingdom
Wyler, S., Ph.D., University of Texas Southwestern, Dallas

SEMINARS

Asahina, K., The Salk Institute for Biological Studies, La Jolla, California: Aggressive interaction as a strategic decision-making process.
Branson, K., Janelia Farm/HHMI, Ashburn, Virginia: Automatic tracking and behavior analysis of model organisms.
Frank, C.A., University of Iowa, Iowa City: Synaptic homeostasis at the *Drosophila* neuromuscular junction.
Gunay, C., Georgia Gwinnett College, Lawrenceville, Georgia: *Drosophila* computer modeling.
Heckscher, E., University of Chicago, Illinois: Development and function of neural circuits in the motor system.
Jayaraman, V., Howard Hughes Medical Institute, Ashburn, Virginia: Abstract internal representations and attractor dynamics in the fly brain.
Kaun, K., Brown University, Providence, Rhode Island: Neurofly bootcamp: Fly biology and genetics 101, fly neuroanatomy and neurogenetic tools, basic microscopy. Designing behavior experiments and building behavior apparatuses. Molecular and circuit mechanisms underlying addiction.
Keene, A., Florida Atlantic University, Jupiter: Neurofly bootcamp: Fly biology and genetics 101, fly neuroanatomy and neurogenetic tools, basic microscopy. Scientific rigor and reproducibility in *Drosophila* neurobiology research. Neural basis of the interaction between sleep and feeding.
Keller, P., Howard Hughes Medical Institute, Ashburn, Virginia: Whole-CNS functional imaging in behaving animals using light-sheet microscopy.
Kohwi, M., Columbia University, New York, New York: Neuroblast development.
Lee, C-H., National Institute of Child Health and Human Development/NIH, Bethesda, Maryland: Neurofly bootcamp: Fly biology and genetics 101, fly neuroanatomy and neurogenetic tools, basic microscopy. Development and function of color-vision circuits in *Drosophila*.
Levine, J., University of Toronto, Mississauga, Ontario, Canada: Genetic basis of social interactions in *Drosophila*.
Louis, M., University of California, Santa Barbara: Orientation behavior in the *Drosophila* larva.
Masek, P., Binghamton University, New York: Gustatory learning and memory.
Muraro, N., University of Manchester, United Kingdom: Electrophysiology in central neurons of *Drosophila*/ circadian rhythms and electrophysiology of clock neurons.
Prinz, A., Emory University, Atlanta, Georgia: Computational modeling of fly neurons.
Pulver, S., University of St. Andrews, United Kingdom: Neurofly bootcamp: Fly biology and genetics 101, fly neuroanatomy and neurogenetic tools, basic microscopy. Neural control of locomotion in *Drosophila* larvae.
Reiser, M., Howard Hughes Medical Institute, Ashburn, Virginia: Connectivity and ultrastructure of the fly eye. Following the visual motion pathway from the retina into the central brain.
Robie, A., Janelia Research Campus/HHMI, Ashburn, Virginia: Automated behavior analysis.
Sehgal, A., The University of Pennsylvania, Philadelphia: Regulation of sleep in *Drosophila* (joint with Advanced Techniques in Molecular Neuroscience Course).
Shao, M., Howard Hughes Medical Institute, Ashburn, Virginia: Synaptic physiology in larval sensory motor circuits.
Wildonger, J., University of Wisconsin, Madison: Use of CRISPR-Cas9 for genome targeting and engineering. Neuronal cell biology.
Zhang, B., University of Missouri, Columbia: Glial, circuits and behavior in *Drosophila*.

Frontiers and Techniques in Plant Science

June 30–July 20

INSTRUCTORS **S. Cutler,** University of California, Riverside
U. Paszkowski, University of Cambridge, United Kingdom
N. Provart, University of Toronto, Ontario, Canada

ASSISTANTS **K. Brown,** University of Kansas, Lawrence
L. Luginbuehl, John Innes Centre, Norwich, United Kingdom
R. Nasti, University of Minnesota, St. Paul

This course provided an intensive overview of topics in plant genetics, physiology, biochemistry, development, and evolution and hands-on experiences in molecular, analytical, computational and high-throughput approaches to understanding plant biology. It emphasized recent results from model organisms including *Arabidopsis,* maize, and tomato, as well as a variety of other plants, and provided an introduction to current methods used in basic and applied plant biology, both theoretically and practically.

The seminar series included plant morphology and anatomy, development, evolution, light and circadian biology, hormones, small RNAs and epigenetic inheritance, biotic and abiotic interactions, plant biochemistry, crop domestication, and applications addressing current agronomic problems. Speakers provided expert overviews of their fields, followed by in-depth discussions of their own work. The laboratory sessions provided exposure to cutting-edge experimental and computational techniques currently used in plant research. These included approaches for studying plant development, transient gene expression, applications of fluorescent proteins, genome editing, and chromatin immunoprecipitation. Students also gained hands-on experience with computational tools and environments for genome assembly, transcriptomics, construction of gene

regulatory networks, identification of quantitative trait loci, mapping by sequencing, and mathematical modeling of development and hormone action. The course also included several short workshops on important themes in plant research. Throughout the course, students interacted individually and informally with the speakers to further enrich the learning experience.

This course was supported with funds provided by the National Science Foundation.

PARTICIPANTS

Araus, V., Ph.D., New York University, New York
Barnhart, M., M.S., SUNY Buffalo, New York
Ellison, E., B.S., University of Minnesota, St. Paul
Gala, H., Ph.D., University of Washington, Seattle
Henderson, A., B.S., University of Kansas, Lawrence
Jaber, A., B.S., Leibniz Institute for Plant Biochemistry, Halle (Saale), Germany
Klesen, S., M.S., University of Tübingen, Germany
Mihelich, N., B.S., University of Minnesota, St. Paul
Palla, K., B.S., University of Tennessee/Oak Ridge National Laboratory, Knoxville
Persson Hodén, K., B.S., Swedish University of Agricultural Sciences, Uppsala, Sweden
Razzaque, S., M.S., The University of Texas, Austin
Sanden, N.C., M.S., Norwegian University of Science and Technology, Trondheim, Norway
Sztojka, B., B.Sc., Umeå Plant Science Centre, Sweden
Tamim, S., B.S., University of Delaware, Newark
Toth, T., B.S., Massachusetts Institute of Technology, Cambridge
Yarahmadov, T., M.S., University of Bern, Switzerland

SEMINARS

Bailey-Serres, J., University of California, Riverside: INTACT and water extremes signaling.
Baulcombe, D., University of Cambridge, United Kingdom: Small RNAs.
Brady, S., University of California, Davis: Gene regulatory networks.
Breakspear, A., John Innes Center, Norwich, United Kingdom: Golden gate cloning.
Cutler, S., University of California, Riverside: Chemical and genetic dissection of plant hormone pathways.
Dinneny, J., Carnegie Institution for Science, Stanford, California: Stressed! How roots cope through dynamic behaviors.
Doebley, J., University of Wisconsin, Madison: Domestication of crops in general and of maize in particular.
Geldner, N., University of Lausanne, Switzerland: Root endodermal polarity.
Hibberd, J., Cambridge University, United Kingdom: Photosynthesis.
Johnson, M., Brown University, Providence, Rhode Island: Molecular dialogs between pollen and Pistil.
Koenig, D., University of California, Riverside: Computational genetics.
Law, J., The Salk Institute, La Jolla, California: Introduction to epigenetics and chromatin immunoprecipitation (ChIP). Epigenetics and DNA methylation.
Michelmore, R., University of California, Davis: Disruptive technologies for understanding and improving disease resistance in crop plants, so many genomes, so little time… .
Nelson, D., University of California, Riverside: Plant hormones.
Oldroyd, G., John Innes Centre, Norwich, United Kingdom: Modulation and symbiosis.
Paszkowski, U., University of Cambridge, United Kingdom: Molecular genetics of *Arbuscular mycorrhizal* symbiosis in cereals.
Peck, S., University of Missouri, Columbia: Introductory talk to analysis of membrane proteins. Analysis of membrane proteins I. Analysis of membrane proteins II. Non-self-recognition between hosts and pathogens.
Pedmale, U., Cold Spring Harbor Laboratory: Light perception and signaling in plants.
Provart, N., University of Toronto, Ontario, Canada: Hypothesis generation with big data.
Shiu, S-H., Michigan State University, East Lansing: Genome evolution.
Sinha, N., University of California, Davis: An introduction to plant biology. Leaves: Connecting form with function.
Surridge, C., SpringerNature, London, United Kingdom: Publishing without tears.
Timmermans, M., University of Tübingen, Germany: Small RNAs as morphogens in development.
Williams, J., Cold Spring Harbor Laboratory: Genomics data carpentry: Basic research skills in R.
Zhang, F., Calyxt, New Brighten, Minnesota: Precise genome engineering with sequence-specific nucleases.

Neural Data Analysis

July 15–28

INSTRUCTORS **M. Reimers,** Michigan State University, East Lansing
P. Wallisch, New York University, New York

ASSISTANTS **M. Moore,** Michigan State University, East Lansing
A. Parthasarathy, Agency for Science, Technology and Research, Singapore
D. Popovkina, University of Washington, Seattle
M. Sarvestani, University of Pennsylvania, Philadelphia
A. Song, Princeton University, New Jersey
M. Valley, Allen Institute for Brain Science, Seattle, Washington

Today's technologies enable neuroscientists to gather data in previously unimagined quantities. This necessitates, and allows for, the development of new analysis methods to address dynamic systems function of brain networks. This course was designed to help neuroscience practitioners develop the conceptual and practical capabilities to meet the challenges posed by the analysis of these hard-won and large data sets. We emphasized statistical issues such as the preprocessing of data, sampling biases, estimation methods, and hypothesis testing as well as data wrangling (in MATLAB). We worked with data from a variety of recording technologies including single- and multi-electrode extracellular recordings, local field potentials, and EEG as well as two-photon and wide-field optical imaging. The course gave a solid conceptual and technical grounding in widely applicable methods such as spectral analysis, multivariate analysis, and network inference, as well as methods specific to each recording technique. We also considered how to integrate neural data with behavioral data.

The workshop proceeded in a seminar style, guided by leading neural data analysts, with demonstrations and practical lab data analysis exercises supervised by instructors.

This course was supported with funds provided by the Helmsley Charitable Trust.

PARTICIPANTS

Abbas, F., Ph.D., Yale University, New Haven, Connecticut
Arriaga, M., B.S., Washington University School of Medicine in St. Louis, Missouri
Auerbach, B., Ph.D., SUNY Buffalo, New York
Dede, A., Ph.D., University of Washington, Seattle
Dooley, J., Ph.D., University of Iowa, Iowa City
Failor, S., Ph.D., University College London, United Kingdom
Gindrat, A-D., M.S./Ph.D., Deutsches Primatenzentrum GmbH, Göttingen, Germany
Hoffmann, M., M.S., Charité University Hospital Berlin, Germany
Hou, H., Ph.D., Harvard University, Boston, Massachusetts
Hsu, C-C., B.S., Emory University, Atlanta, Georgia
Huang, B., Ph.D., Weill Cornell Medical College, New York, New York
Li, S. (William), B.S., Massachusetts General Hospital, Boston
Murata, Y., Ph.D., George Washington University, Washington, D.C.
Patino, A., B.S., Novartis Institutes for Biomedical Research, Cambridge, Massachusetts
Ramaswamy, M., B.Tech., National University of Singapore, Singapore
Sahraee Ardakan, M., B.S., University of California, Los Angeles
Skelin, I., Ph.D., University of Lethbridge, Alberta, Canada
Swanson, R., M.S., New York University, New York
Yeh, C-M., Ph.D., The Salk Institute for Biological Studies, San Diego, California
Zemla, R., B.A., New York University School of Medicine, New York

SEMINARS

Babadi, B., University of Maryland, College Park: Methods and limitations of inferring networks from spiking data. Inferring networks from calcium data (practical).
Bassett, D., University of Pennsylvania, Philadelphia: Using graph theory to understand interaction patterns in neural data.
Churchland, A., Cold Spring Harbor Laboratory: International brain laboratory.
Cohen, M., University of Amsterdam, Netherlands: Analog signals I (LFP); Time and frequency domain analysis. Spike sorting: An unsolvable problem. Analog signals II (LFP): Linear algebra + least squares. Analog signals II (LFP): Eigen decomposition and source separation. Analog signals II (LFP): Space curves.
Giovannucci, A., Simons Center for Data Analysis, New York, New York: CaImAn: An open source toolbox for large-scale calcium imaging data analysis on standalone machines.
Iyer, V., MathWorks, Natick, Massachusetts: New parallel computing MATLAB functions for big data in neuroscience.
Kleinfeld, D., University of California, San Diego, La Jolla: Electrons, photons, and viruses to reverse engineer sensorimotor dynamics, Part one. Electrons, photons, and viruses to reverse engineer sensorimotor dynamics, Part two. The good, the bad, and the ugly of emerging directions in neurotechnology.
Kording, K., Northwestern University, Chicago, Illinois: How to write a paper. Intro to spike analysis and machine learning. Approaches to decoding population data: Classification methods. Debate: How to make neuroscience better.
Mohajerani, M., University of Lethbridge, Canada: Wide-field optical imaging: Characteristics of different imaging modalities and indicators. Research questions that can be addressed with wide-field optical methods. Quantitative analysis toolbox for characterization of spatiotemporal dynamics in mesoscale optical imaging of brain activity.
Pachitariu, M., University College London, United Kingdom: What you can find out by recording 10K neurons. Finding cells in calcium images: The Suite2P approach. Using Suite2P (practical). High-density electrophysiology: Spike sorting and analyzing neuropixel data.
Pillow, J., Princeton University, New Jersey: Spikes II: ENCODING—Model-based spike analysis. Encoding models using GLM (exercise).
Pnevmatikakis, E., Simons Center for Data Analysis, Simons Foundation, New York, New York: CaImAn: An open source toolbox for large-scale calcium imaging data analysis on standalone machines.
Reimers, M., Michigan State University, East Lansing: Introduction and approach to neural data analysis. Debate: How to make neuroscience better. The landscape of dimension reduction methods. Visualizing state space: Dimensionality reduction method. Synthesis.
Smith, M., University of Pittsburgh, Pennsylvania: Spikes III: Array analysis (spike count correlations).
Wallisch, P., New York University, New York: Introduction and approach to neural data analysis. Analysis of "big" neural data. Intro to MATLAB: Exploring the data set. Of mice, monkeys, and (wo)men. Synthesis.

Synthetic Biology

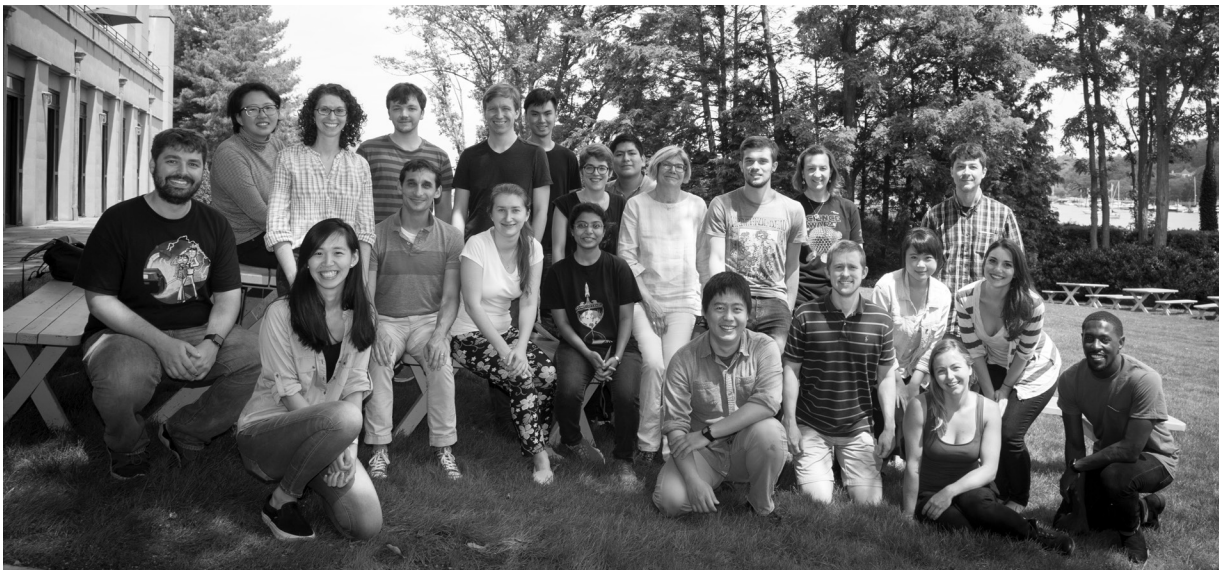
July 25–August 7

INSTRUCTORS
C. Beisel, North Carolina State University, Raleigh
V. Noireaux, University of Minnesota, Minneapolis
H. Salis, Pennsylvania State University, State College
L. Woodruff, University of Massachusetts, Amherst

ASSISTANTS
S. Collins, North Carolina State University, Raleigh
C. Cuba Samaniego, University of California, Riverside
S. Halper, Pennsylvania State University, State College
N. Howitz, University of Massachusetts, Amherst
R. Marshall, University of Minnesota, Minneapolis

Synthetic biology is a discipline in which living organisms are genetically programmed to carry out desired functions in a reliable manner. The field takes inspiration from our ever-expanding ability to measure and manipulate biological systems and also from the philosophical reflections of Schrödinger and Feynman—specifically, that physical laws can be used to describe and rationally engineer biology to accomplish useful goals. Cells are the world's most sophisticated chemists, and their ability to adapt to changing environments offers enormous potential for solving modern engineering challenges. Nonetheless, biological systems are noisy, massively interconnected, and nonlinear, and they have not evolved to be easily engineered. The grand challenge of synthetic biology is to reconcile the desire for a predictable, formalized biological design process with the inherent "squishiness" of biology.

This course focused on how the complexity of biological systems can be combined with traditional engineering approaches to result in new design principles for synthetic biology. The



centerpiece of the course was an immersive laboratory experience in which students worked in teams to learn the practical and theoretical underpinnings of synthetic biology research. Broadly, the course explored how cellular regulation (transcriptional, translational, posttranslational, and epigenetic) can be used to engineer cells that accomplish well defined goals. Laboratory modules covered the following areas: CRISPR technologies for genome editing and gene regulation; cell-free transcription and translation systems to characterize genetic circuits and RNA regulators; modeling gene expression using ordinary differential equations; and high-throughput DNA assembly techniques and genetic design principles.

Students first learned essential synthetic biology techniques in a four-day "boot camp" at the beginning of the course. Following the boot camp, they rotated through research projects in select areas. Students also interacted closely with a panel of internationally recognized speakers who collectively provided a broad overview of synthetic biology applications, including renewable chemical production and therapeutics, state-of-the-art techniques, case studies in human practices, and socially responsible innovation.

This course was supported with funds provided by the National Institute of General Medical Sciences, Howard Hughes Medical Institute, Helmsley Charitable Trust, and National Science Foundation.

PARTICIPANTS

Dalhuijsen, S., M.S., DSM Biotechnology Center, Delft, Netherlands
Donis-Keller, H., Ph.D., F.W. Olin College of Engineering, Needham, Massachusetts
Doshi, A., M.Sc., University of Houston, Texas
Green, L., Ph.D., California Institute of Technology, Pasadena
Hikiji Watanabe, L., B.S., University of Utah, Salt Lake City
Kan, J., Ph.D./M.A., California Institute of Technology, Pasadena
Korniy, N., B.S., Max Planck Institute for Biophysical Chemistry, Göttingen, Germany
Malico, A., B.A., North Carolina State University, Raleigh
Rosa, L., M.Sc., University College London, United Kingdom
Ross, C., B.S., Arizona State University, Tempe
Rostain, W., Ph.D., University of Warwick, Coventry, United Kingdom
Song, F., Ph.D., Lawrence Berkeley National Lab, Berkeley, California
Sycuro, L., Ph.D., University of Calgary, Alberta, Canada
Tsai, Y-T., Ph.D., Institute of Molecular and Cell Biology of Rosario, Argentina
Valet, M., M.S., Laboratoire Jean Perrin UMR 8237-CNRS-UPMC, Paris, France
Wu, F., B.S., Duke University, Durham, North Carolina

SEMINARS

Del Vecchio, D., Massachusetts Institute of Technology, Cambridge: Context-dependence and its mitigation in synthetic genetic circuits.
Dueber, J., University of California, Berkeley, San Francisco: Synthetic biology strategies for gaining spatial control over engineered metabolic pathways (joint lecture with Yeast Genetics and Genomics course).
Endy, D., Stanford University, Menlo Park, California: What should we be arguing about?
Glass, J., J. Craig Venter Institute, La Jolla, California: Design, construction, and analysis of a synthetic minimal bacterial cell.
Hoff, K., Genomatica, Inc., San Diego, California: Engineering microorganisms for commercial success.
Li, S., Stanford University, California: Biosynthesis of noscapine and derivatives in yeast.
Romero, P., University of Wisconsin, Madison: Data-driven protein engineering.
Venturelli, O., University of Wisconsin, Madison: Deciphering microbial interactions that shape the dynamics of the human gut microbiota.
Wang, H., Columbia University, New York, New York: Engineering microbial communities through mobile trans-species gene circuits.

Chromatin, Epigenetics, and Gene Expression

July 25–August 14

INSTRUCTORS **K. Adelman,** Harvard Medical School, Boston, Massachusetts
L. Di Croce, ICREA and Centre for Genomic Regulation, Barcelona, Spain
G. Narlikar, University of California, San Francisco
A. Shilatifard, Northwestern University Feinberg School of Medicine, Chicago, Illinois

ASSISTANTS **N. Gamarra,** University of California, San Francisco
T. Henriques, Harvard Medical School, Boston, Massachusetts
M. Keenen, University of California, San Francisco
F. Le Dily, Centre for Genomic Regulation, Barcelona, Spain
K. (Kevin) Liang, Northwestern University, Chicago, Illinois
M. Morgan, Northwestern University Feinberg School of Medicine, Chicago, Illinois

This course was designed for students, postdocs, and principal investigators who have recently ventured into the exciting area of gene regulation. Emphasis was placed on exposing students to a broad array of methodologies to study gene regulation, chromatin structure, and dynamics, including both state-of-the-art and well developed methods.

Students performed widely used techniques such as chromatin immunoprecipitation (ChIP) coupled with sequencing (ChIP-seq), reporter assays of enhancer activity, and RNA analysis. They applied a basic pipeline to analyze sequencing results and discussed current informatics strategies. Students isolated transcription factor complexes and assessed their activity in functional assays, knocked down specific factors using RNAi, and evaluated the effects on gene expression.

This course provided the basic concepts behind different methods used to analyze the chromatin architecture of the genome. Students performed chromosome conformation capture (3C)

experiments, together with other approaches aimed at interrogating the three-dimensional organization of genomes. Moreover, we discussed the computational methods required to analyze these data.

Students learned how to assemble recombinant chromatin and use biophysical methods such as FRET to assay the activity of chromatin remodeling enzymes. They also learned principles of enzyme kinetics and applied these to quantify the remodeling reactions.

Given the broad biological roles for DNA-binding transcription factors, and emerging roles of noncoding RNAs in transcription regulation, electrophoretic mobility shift assays (EMSAs) are again becoming widely used for assessing transcription factor binding to regulatory DNA or RNA elements. Students learned how to perform and interpret EMSA experiments, using both microscale thermophoresis and gel-based methods.

Experience with basic recombinant DNA techniques was a prerequisite for admission to this course. Lectures by the instructors covered the current state of the gene expression field, theoretical aspects of the methodology, and broader issues regarding strategies for investigating the regulation of gene expression in eukaryotes. Emphasis was placed on advantages and limitations of specific techniques and data interpretation. Each evening, an invited speaker who is an expert in the field presented their work and interacted with students. The students were encouraged and expected to actively participate in these discussions and to take advantage of the many opportunities to network and receive input on their projects and future plans.

This course was supported with funds provided by the National Cancer Institute.

PARTICIPANTS

Boehm, U., Ph.D., National Cancer Institute/NIH, Bethesda, Maryland
Buchacher, T., Ph.D., University of Turku, Finland
Cai, W., B.S., Yale School of Medicine, New Haven, Connecticut
Duke, C., B.S., University of Alabama, Birmingham
Evans, S., B.A., Brown University, Providence, Rhode Island
Haro Acosta, V., B.S., University of Southern California, Los Angeles
Himanen, S., M.S., Åbo Akademi University, Turku, Finland
Horwitz, G., Ph.D., University of Washington, Seattle
Leonidou, A., Ph.D., The Institute of Cancer Research, London, United Kingdom
Lim, J., B.Sc., Queensland Brain Institute, St Lucia, Queensland, Australia
Mariossi, A., M.S., King's College London, United Kingdom
Philips, R., B.S., Mayo Clinic Graduate School of Biomedical Sciences, Rochester, Minnesota
Ray, A., M.Sc., University of Illinois, Urbana-Champaign
Rudigier, L., Ph.D., Humboldt-Universität zu Berlin, Germany
Vaziri, A., B.S., University of Michigan, Ann Arbor
Wu, X., B.S., Stony Brook University, New York

SEMINARS

Bonasio, R., University of Pennsylvania, Philadelphia: RNA-mediated regulation of chromatin and epigenetics.
Farley, E., University of California, San Diego, La Jolla: Regulatory principles governing enhancer function during development.
Fuks, F., University of Brussels, Germany: DNA and RNA modifications.
Johnson, T., University of California, Los Angeles: Chromatin modification and RNA processing: Unraveling the ties that bind.
Kadonaga, J., University of California, San Diego, La Jolla: Operating systems, apps, and a novel chromatin structure for the regulation of our genes.
Kingston, R., Massachusetts General Hospital/Harvard Medical School, Boston: Regulation of gene expression during development by the Polycomb group.
Lambowitz, A., University of Texas, Austin: How mobile group ii introns shaped the human genome and enable new RNA-Seq methods for gene expression analysis.
Levine, M., Princeton University, New Jersey: Regulation of enhancer–promoter interactions in living *Drosophila* embryos.
Lis, J., Cornell University, Ithaca, New York: Dissecting molecular mechanisms of transcription and its regulation in living cells.

Luger, K., University of Colorado, Boulder: 20 years of the nucleosome: Something old, something new, something borrowed, and something blue.

Marti-Renom, M., CNAG-CRG, Barcelona, Spain: Structure determination of genomes and genomic domains by satisfaction of spatial restraint.

Mendillo, M., Northwestern University School of Medicine, Chicago, Illinois: The left and right leanings of heat shock factor 1.

Rao, A., La Jolla Institute for Allergy & Immunology, California: TET methylcytosine oxidases, immune responses, and cancer.

Roeder, R., The Rockefeller University, New York: Elucidation of eukaryotic transcriptional regulatory mechanisms through biochemical approaches.

Schneider, R., HMGU, Münich, Germany: Novel players in chromatin.

Stark, A., Research Institute of Molecular Pathology, Vienna, Austria: Decoding transcriptional regulation.

Torres-Padilla, M., Institute of Epigenetics & Stem Cells, Münich, Germany: Epigenetic regulation of early development.

Witkowski, J., Cold Spring Harbor Laboratory: Ethics, reproducibility, and vigor.

Imaging Structure and Function in the Nervous System

July 25–August 14

INSTRUCTORS
F. Albeanu, Cold Spring Harbor Laboratory
M. Orger, Champalimaud Foundation, Lisbon, Portugal
L. Palmer, University of Melbourne, Victoria, Australia
P. Tsai, University of California, San Diego

ASSISTANTS
A. Bandyopadhyay, New York University Langone Medical Center, New York
J. Donovan, Max-Planck Institute of Neurobiology, Martinsried, Germany
C. Grienberger, Janelia Research Campus/HHMI, Ashburn, Virginia
S. Renninger, Champalimaud Foundation, Lisbon, Portugal
S. Schröder, University College London, United Kingdom
N. Takahashi, Humboldt University of Berlin, Germany
D. Wilson, Max Planck Florida Institute for Neuroscience, Jupiter

Advances in light microscopy and digital image processing and the development of a variety of powerful fluorescent probes present expanding opportunities for investigating the nervous system, from synaptic spines to networks in the brain. This intensive laboratory and lecture course provided participants with the theoretical and practical knowledge to use emergent imaging technologies. The primary emphasis of the course was on vital light microscopy. Students learned the principles of light microscopy, as well as the use of different types of cameras, laser-scanning systems, functional fluorophores, delivery techniques, and digital image-processing software. In addition to transmitted light microscopy for viewing cellular structure, the course examined a variety of molecular

probes of cell function, including calcium-sensitive dyes and optogenetic and photo-activatable molecules. Particular weight was given to multiphoton laser-scanning microscopy and to biological fluorophores, especially green fluorescent protein (GFP) and its variants. We used a range of neural and cell biological systems, including living animals, brain slices, and cultured cells.

This course was supported with funds provided by the Helmsley Charitable Trust and Howard Hughes Medical Institute.

PARTICIPANTS

d'Aquin, S., M.S., Friedrich Miescher Institute for Biomedical Research, Basel, Switzerland
de Oliveira Rodrigues Vaz da Cunha, I., M.S., Columbia University, New York, New York
Gilday, O., B.Sc., The Hebrew University in Jerusalem, Israel
Hamid, A., Ph.D., Brown University, Providence, Rhode Island
Han, S., B.S., Columbia University, New York, New York
Laboy-Juárez, K., Ph.D., University of California, Berkeley
Lopez Ortega, E., B.S., The Johns Hopkins University School of Medicine, Baltimore, Maryland
Nashaat, M., Ph.D., Humboldt Universität zu Berlin, Germany
Palumbo, F., M.S., NTNU-Faculty of Medicine, Trondheim, Norway
Sehgal, M., M.S./Ph.D., University of California, Los Angeles
Sotuyo, N., B.S./M.S., University of Pennsylvania VMD/PhD Program, Children's Hospital of Philadelphia
Sundararajan, J., B.Eng., Duke University, Durham, North Carolina
Wixted, E., B.S., University of Michigan, Ann Arbor
Zhang, Z., B.Eng., University College London, United Kingdom

SEMINARS

Albeanu, F., Cold Spring Harbor Laboratory: Widefield/intrinsic. Intro to building a homebrew 2P microscope including laser safety.
Deisseroth, K., Stanford University, California: Optogenetics: Recent advances/clarity.
Denk, W., Max-Planck Institute of Neurobiology, Martinsried, Germany: Block-face EM/connectomics.
Dieudonné, S., Ecole Normale Supérieure, Paris, France: AODs and random access multiphoton imaging.
Dombeck, D., Northwestern University, Evanston, Illinois: Awake imaging in hippocampus.
Emiliani, V., CNRS and University Paris Decartes, France: Holographic photoactivation. SLMs and holography.
Engert, F., Harvard University, Cambridge, Massachusetts: From functional and structural data to circuit understanding.
Hillman, E., Columbia University, New York, New York: High-speed volume imaging with SCAPE microscopy.
Holy, T., Washington University, Saint Louis, Missouri: Light-sheet microscopy.
Ji, N., University of California, Berkeley: Super-resolution imaging: Structured illumination, PALM, STED. Deep imaging/adaptive optics.
Judkewitz, B., Charité and Humboldt University, Berlin, Germany: 2P microscopy; lasers. Applications of 2P microscopy.
Lichtman, J., Harvard University, Cambridge, Massachusetts: Principles and practice of confocal microscopy. Applications of confocal and other imaging approaches to connectomics.
Lin, M., Stanford University, California: Genetically encoded indicators of neuronal activity.
Looger, L., Janelia Research Campus/HHMI, Ashburn, Virginia: Latest developments in molecular tools.
Narasimhan, A., Cold Spring Harbor Laboratory: Clearing brains.
Oertner, T., Hamburg University, Hamburg, Germany: Intro to optogenetics.
Orger, M., Champalimaud Foundation, Lisbon, Portugal: Intro: Scanning/PMTs.
Peterka, D., Columbia University, New York, New York: Incoherent emission point source, PSF/Airy pattern, Fourier/OTF, aberrations. Phase contrast and DIC, Fourier/OTF, introducing SLMs.
Tsai, P., University of California, San Diego: Nature of light, coherence, wave description, polarization. Ray tracing, lenses. Introduction to Optics Bench Lab. Huygens principle, law of refraction, diffraction, Abbe resolution limit. Kohler illumination, aperture, and field control.
Vaziri, A., The Rockefeller University, New York, New York; Sculpted excitation for imaging and optogenetics.
Waters, J., Allen Institute for Brain Science, Seattle, Washington: In vivo imaging in rodents/analyzing calcium imaging data.
Waters, J., Harvard Medical School, Boston, Massachusetts: Noise and detectors.
Witten, I., Princeton University, New Jersey: Fiber photometry.
Xu, C., Cornell University, New York, New York: 3P imaging.

Yeast Genetics and Genomics

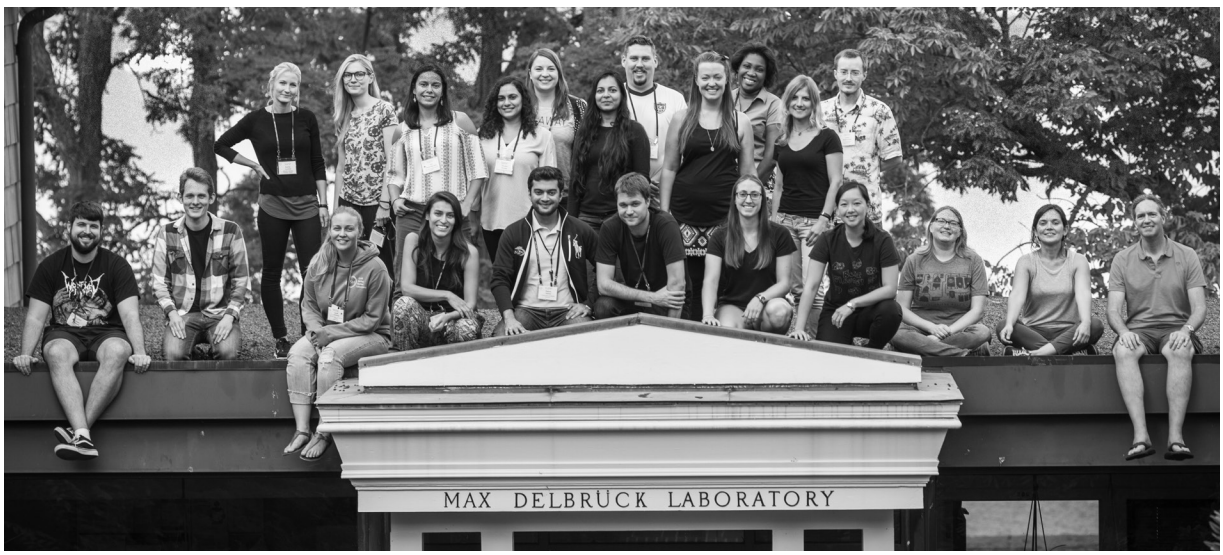
July 25–August 14

INSTRUCTORS **G. Brown,** University of Toronto, Ontario, Canada
M. Dunham, University of Washington, Seattle
E. Ünal, University of California, Berkeley

ASSISTANTS **C. Amorosi,** University of Washington, Seattle
E. Sawyer, University of California, Berkeley
T. Sing, University of Toronto, Ontario, Canada

This course was a modern, state-of-the-art, intensive laboratory course designed to teach students the full repertoire of genetic approaches needed to dissect complex problems in the yeast *Saccharomyces cerevisiae*. Combinations of classical and modern genetic approaches were emphasized, including the isolation and characterization of mutants, tetrad analysis, and complementation. Synthetic biology was explored through CRISPR-Cas9-directed engineering of heterologous biosynthetic pathways in yeast. Students learned genome-based methods of analysis facilitated by the yeast genome sequence, the gene deletion collection, and other genomic resources available to the community. Molecular genetic techniques, including yeast transformation, gene replacement by PCR, construction and analysis of gene fusions, and generation of mutations, were also emphasized.

Students used classical approaches and modern whole-genome sequencing to gain experience in identifying and interpreting different kinds of genetic interactions, including suppression, synthetic lethality, and chemical–genetic interactions. They performed genome-scale screens using the synthetic genetic array (SGA) methodology and were immersed in yeast genomics and performed and interpreted experiments using DNA arrays, colony arrays, whole-genome sequencing,



and multiplexed DNA barcode sequencing. Computational methods for data analysis were introduced. Students gained firsthand experience in modern cytological approaches such as epitope tagging and imaging yeast cells using fluorescence microscopy with GFP-protein fusions and fluorescent indicators for different subcellular structures and organelles. Lectures on fundamental aspects of yeast genetics were presented along with seminars given by prominent experts in the field on topics of current interest.

This course was supported with funds provided by the National Science Foundation.

PARTICIPANTS

Aguilar-Rodriguez, J., Ph.D., University of Zürich, Switzerland
Boocock, J., Ph.D., University of California, Los Angeles
Dhami, M., Ph.D., Stanford University, California
D'Orazio, K., B.S./M.S., The Johns Hopkins University School of Medicine, Baltimore, Maryland
Flores-Bojorquez, A., M.S., University of California, San Diego, La Jolla
Kutch, I., Ph.D., Oregon State University, Corvallis
Langstein, I., M.S., Ludwig-Maximilian University of Münich, Planegg-Martinsried, Germany
Løvschal, K., M.S., University of Copenhagen, Denmark
McKerracher, A., B.S., AB Mauri Food, Inc., Saint Louis, Missouri
Nivedita, N., Ph.D., University of North Carolina, Chapel Hill
Sachsenhauser, V., M.Sc., University of Michigan, Ann Arbor
Samani, P., M.S./Ph.D., Georgia Institute of Technology, Atlanta
Thomas, A., Ph.D., Crown College, St. Bonifacius, Minnesota
van Dijk, M., M.Sc., Chalmers University of Technology, Göteborg, Sweden
Vydzhak, O., M.Sc., Institute of Molecular Biology, Mainz, Germany
Zuniga, A., Ph.D., Centre de Biochimie Structurale, Montpellier, France

SEMINARS

Baryshnikova, A., Calico Life Sciences, South San Francisco, California: Systematic phenotypic analysis of single and double perturbations in yeast.
Biggins, S., Fred Hutchinson Cancer Research Center, Seattle, Washington: Reconstituting kinetochore functions in vitro.
Botstein, D., Calico Labs, South San Francisco, California: Understanding cellular stress response at the system level in yeast.
Brar, G., University of California, Berkeley: Decoding gene expression regulatory principles in meiosis.
Cowen, L., University of Toronto, Ontario, Canada: Functional genomic analysis of fungal morphogenesis and immune evasion.
Drummond, A., The University of Chicago, Illinois: Rethinking yeast heat-shock responses.
Dueber, J., University of California, Berkeley, San Francisco: Synthetic biology strategies for gaining spatial control over engineered metabolic pathways (joint seminar with the Statistical Methods for Functional Genomics course).
Fortmann, K., White Labs, San Diego, California: Omics on tap: Domestication and divergence of *Saccharomyces cerevisiae* beer yeasts.
Gartenberg, M., Robert Wood Johnson Medical School, Piscataway, New Jersey: Heterochromatin, euchromatin, and the cohesion of sister chromatids.
Klein, H., New York University School of Medicine, New York: Genome integrity and ribonucleotides in DNA.
Lacefield, S., Indiana University, Bloomington: The power of live-cell microscopy to assess variability in individual cells.
Lang, G., Lehigh University, Bethlehem, Pennsylvania: Genome evolution in laboratory populations.
Madhani, H., University of California, San Francisco: Epigenetic memory over geological timescales.
Nash, R., Stanford University, Palo Alto, California: Navigating data at SGD with YeastMine.
Schacherer, J., University of Strasbourg, Alabama: From genotype to phenotype: Insight from species-wide surveys.

AMAR KLAR MEMORIAL

We were all stunned to learn in early 2017 of the tragic death of Amar Klar, a beloved former member of Cold Spring Harbor Laboratory. Amar was one of the trio, with Jim Hicks and Jeff Strathern, that made up the famous Yeast Group that flourished at CSHL in the early 1980s. So, there was no

hesitation in accepting the suggestion of the organizers that the Laboratory was the place to hold a Symposium in conjunction with the yeast course to celebrate Amar's life.

It was in 1978 that the Yeast Group came into being when Watson hired first Jeff Strathern, then Jim Hicks, and finally Amar Klar. They set up base in Davenport Laboratory, and despite the primitive conditions—Demerec was much in need of renovation—they set to work to determine the molecular basis of mating-type switching in yeast. They succeeded and published a series of classic papers between 1979 and 1984. The Yeast Group came to a gradual end, Strathern leaving CSHL in 1984 and Hicks in 1986. Amar moved to the National Cancer Institute, Frederick, in 1988, but he returned frequently to the Laboratory, as irrepressible as ever, telling everyone about his latest ideas.

The esteem in which Amar's work is held and our affection for him were evident at the meeting. Many of Amar's former colleagues joined us, and we were particularly delighted that Amar's family could come and help us celebrate Amar's life.

Workshop on Autism Spectrum Disorders

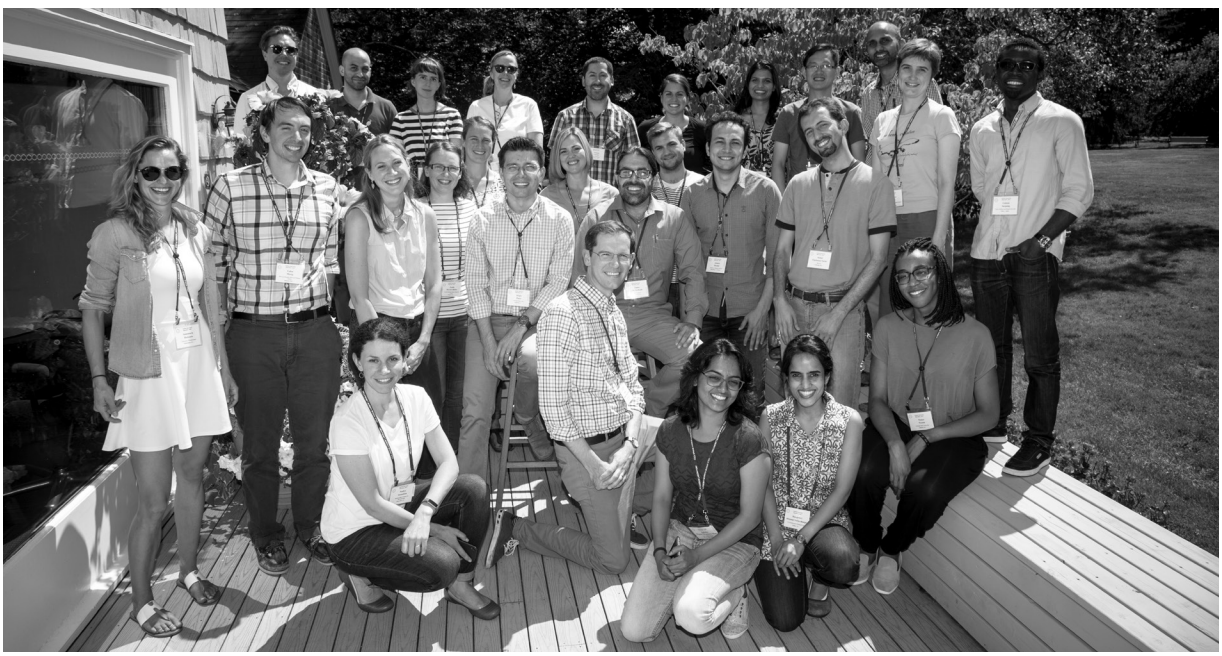
July 31–August 6

INSTRUCTORS **J. McPartland,** Yale University, New Haven, Connecticut
S. Pasca, Stanford University, Stanford, California
J. Veenstra-Vander Weele, Columbia University, New York, New York

ASSISTANT **S. Jackson,** Yale University, New Haven, Connecticut

Autism spectrum disorders (ASDs) are developmental disorders with complex phenotypes defined by a triad of symptoms that include disrupted social abilities, verbal and nonverbal communication skills, and restricted interests with repetitive behaviors. Co-occurring neurological and medical conditions often arise in this disorder. The underlying etiology remains a mystery, but ASDs are some of the most highly heritable of neuropsychiatric disorders. This workshop examined dimensions of ASDs on various levels, including sessions on characteristics of the clinical syndrome; the neuropathology, imaging, and cognitive neuroscience studies that implicate circuits and systems involved in ASDs; the current state of findings from human genetics; concepts regarding the developmental neurobiological basis; the use of experimental models; and current etiological theories and hypotheses of ASDs.

In addition to learning about the most recent research in these areas, we explored and debated controversial topics and challenges of basic assumptions in the field. An exceptional faculty with diverse interests brought the most up-to-date results and theories to the students, making this workshop a valuable resource for young researchers starting out in this fast-moving and expansive field. It not only helped them build the foundation for their future research, but it also introduced



them to many potential collaborators working to understand ASD from different disciplines. The workshop had hands-on exercises to complement the featured intense lecture sessions. Most importantly, students also had free time for reading, informal discussions, and recreation on the beautiful campus of the Banbury Center, which includes a beach, a pool, and tennis courts.

This course was supported with funds provided by the Nancy Lurie Marks Family Foundation.

PARTICIPANTS

Bohorquez, D., Ph.D., Duke University, Durham, North Carolina
Cancino, G., Ph.D., Universidad Mayor, Santiago, Chile
Davis, R., Ph.D., Bibkbeck, University of London, United Kingdom
Eltokhi, A., M.A., Institute of Human Genetics, Heidelberg, Germany
Holehonnur Sudarshanprasad, R., Ph.D., University of Texas Southwestern Medical Center, Dallas
Huttunen, A., Ph.D., University of Pennsylvania, Philadelphia
Kovacs-Balint, Z., Ph.D., Yerkes National Primate Research Center, Atlanta, Georgia
Li, X., Ph.D., McGovern Institute for Brain Research/ Massachusetts Institute of Technology, Cambridge
Lord, K., Ph.D., University of Massachusetts Medical School, Worcester
Missig, G., Ph.D., McLean Hospital, Harvard Medical School, Belmont, Massachusetts
Nabel, E., B.A., Icahn School of Medicine at Mount Sinai, New York
Nwosu, N., B.S., Georgia State University, Atlanta
Pakdaman Naeini, M., Ph.D., Harvard University, Cambridge, Massachusetts
Post, K., B.A., University of British Columbia, Vancouver, Canada
Sarpong, G., B.Sc., Tokyo Medical and Dental University, Japan
Sharon, G., Ph.D., California Institute of Technology, Pasadena
Siddappa Niranjana Murthy, A., M.Sc., University of Mysore, Karnataka, India
Sosonkina, N., Ph.D., HudsonAlpha Institute for Biotechnology, Huntsville, Alabama
Sullivan, C., Ph.D., University of North Carolina, Chapel Hill
Tatavarty, V., Ph.D., Brandeis University, Waltham, Massachusetts
Tora, D., M.A., University of Basel, Switzerland
Yadav, S., Ph.D., University of California, San Francisco

SEMINARS

Champagne, F., Columbia University, New York, New York: Epigenetics in ASD.
Dawson, G., Duke University Medical Center, Durham, North Carolina: Psychological models of ASD.
Ecker, C., University Hospital, Goethe University, Frankfurt, Germany: Neuroimaging methods and findings, Part 1. Neuroimaging methods and findings, Part 2.
Fombonne, E., Oregon Health & Science University, Portland: Epidemiology of ASD. Environmental risk factors.
Geschwind, D., David Geffen School of Medicine, University of California, Los Angeles: Introduction to human genetics and genetic findings in ASD, Part 1. Genetic findings in ASD, Part 2.
Hensch, T., Harvard University, Cambridge, Massachusetts: Critical periods/plasticity mechanisms.
Kasari, C., University of California, Los Angeles: Outcome measures. Behavioral treatments.
Lord, C., Weill Cornell Medicine, White Plains, New York: Historical perspective and clinical presentations.
McPartland, J., Yale University, New Haven, Connecticut: Electrophysiology methods and findings.
Pasca, S., Stanford University, California: Cellular models of disease (iPSC).
Platt, M., University of Pennsylvania, Philadelphia: Primate models of ASD.
Powell, C., University of Texas Southwestern Medical Center, Dallas: Mouse behavior: General. Mouse behavior: Electrophysiology and rescue.
Sabatini, B., Harvard Medical School, Boston, Massachusetts: Synaptic function and dysfunction in ASD.
Sahin, M., Harvard Medical School/Boston Children's Hospital, Boston, Massachusetts: Tuberous sclerosis and related syndromes associated with ASD.
Schumann, C., University of California Davis MIND Institute, Davis: Neuropathology challenges and findings.
Sestan, N., Yale University Medical School, New Haven, Connecticut: Development of the nervous system in ASD, Part 1. Development of the nervous system in ASD, Part 2.
Singer, A., Autism Science Foundation, New York: Parent/ sibling perspective.
Spence, S., Children's Hospital Boston/Harvard University, Boston, Massachusetts: Neurological/medical perspective. Discussion of clinical assessments.
Stevens, B., Children's Hospital Boston/Harvard Medical School, Boston, Massachusetts: Glia in health and disease.
Veenstra Vander Weele, J., Columbia University, New York, New York: Pharmacology: Bench to bedside.

Proteomics

August 1–14

INSTRUCTORS **G. Knudsen,** University of California, San Francisco
D. Pappin, Cold Spring Harbor Laboratory

PART-TIME INSTRUCTORS **E. Soderblom,** Duke University, Durham, North Carolina
W. Thompson, Duke Center for Genomic and Computational Biology, Durham, North Carolina

ASSISTANTS **A. Greninger,** Fred Hutchinson Cancer Research Center, Seattle, Washington
E. Klement, Biological Research Centre of the Hungarian Academy of Science, Szeged, Hungary
A. Makarenko, Cold Spring Harbor Laboratory
K. Rivera, Cold Spring Harbor Laboratory
C. Shaw, Cold Spring Harbor Laboratory
J. Wilson, Cold Spring Harbor Laboratory

This intensive laboratory and lecture course focused on cutting-edge proteomic approaches and technologies. Students gained practical experience isolating, purifying, and identifying protein complexes: Sample preparation with in-solution digestion or FASP protocol was performed, and students were then trained using high-sensitivity nano-HPLC coupled with nanospray-ESI and tandem mass spectrometry analysis. Different search engines and bioinformatics approaches were introduced for data evaluation. Students were shown how to recognize unexpected posttranslational modifications. They used affinity chromatography for phosphopeptide enrichment, and the characterization of the resulting complex mixture, including site assignment, was performed. For shotgun proteomic analysis sections, students used label-free and covalent isotopic-labeling quantitative approaches to profile changes in protein complexes and whole proteomes. Students learned both single-dimension and multidimensional separation methods. In a section focused on

targeted proteomics, students learned to analyze and process shotgun proteomic data for the development of SRM/MRM assays that accurately identify and quantify targeted proteins. Students designed transitions for selected peptides and performed SRM/MRM analyses. They learned to process and interpret the acquired data to measure changing quantities of targeted proteins in a variety of biological samples. For all sections of the course, a strong emphasis was placed on data analysis. There was opportunity to discuss individual research projects.

A series of outside lecturers discussed various proteomics topics including differential gel electrophoresis (DIGE), de novo sequence analysis, advanced mass spectrometry methods, glycosylation, and functional proteomics. The aim of the course was to provide each student with the fundamental knowledge and hands-on experience necessary for performing and analyzing proteomic experiments. The overall goal was to train students to identify new opportunities and applications for proteomic approaches in their biological research.

This course was supported with funds provided by the National Institute of Child Health and Human Development.

PARTICIPANTS

Bayarri-Olmos, R., M.S., Rigshospitalet, University of Copenhagen, Denmark
Casas Garcia, G.P., M.A., La Trobe University, Bundoora, Victoria, Australia
Chigorimbo-Tsikiwa, N., Ph.D., University of Cape Town, South Africa
Dolata, K., M.S., University of Greifswald, Germany
Fleites, L., Ph.D., Boyce Thompson Institute, Ithaca, New York
Hung, V., Ph.D., Stanford University, California
Khudyakov, J., Ph.D., University of the Pacific, Stockton, California
Lapehn, S., B.A., University of Michigan, Ann Arbor
Lobel, L., Ph.D., Harvard T.H. Chan School of Public Health, Boston, Massachusetts
Nevitt, C., B.S., University of Louisville, Kentucky
Qayyum, S., Ph.D., Aligarh Muslim University, India
Rieder, L., Ph.D., Brown University, Providence, Rhode Island
Sandoz, K., Ph.D., Rocky Mountain Labs/NIAID/NIH, Hamilton, Montana
Sinha, A., B.Eng., Nanyang Technological University, Singapore
Soll, J., B.S., Washington University, St. Louis, Missouri
Varadarajan, S., B.Tech., University of Michigan, Ann Arbor
Vincek, A., Ph.D., Mount Sinai School of Medicine, New York

SEMINARS

Anderson, L., NHMFL/Florida State University, Tallahassee: Intact protein characterization.
Clauser, K., Broad Institute of MIT/Harvard University, Cambridge, Massachusetts: Manual de novo peptide MS/MS interpretation for evaluating database search results.
Farnsworth, C., Cell Signaling Technology, Danvers, Massachusetts: PTMScan® technology: An antibody-based proteomics discovery platform.
Heck, M., Cornell University, Ithaca, New York: Understanding how insects transmit pathogens: Lessons from proteomics.
Hendricks, A., AstraZeneca, Waltham, Massachusetts: Introduction of inlysate chemoproteomics and work flow. Industrial case study.
Jacob, R., Matrix Science Inc., Boston, Massachusetts: MS/MS-based protein identifications and the fine-tuning of the search engines.
Kawatkar, A., AstraZeneca, Waltham, Massachusetts: Introduction of inlysate chemoproteomics and work flow. Industrial case study.
Klement, E., Biological Research Centre of the Hungarian Academy of Science, Szeged, Hungary: Phosphopeptide enrichment using metal-ion affinity chromatography.
Knudsen, G., University of California, San Francisco: On the isolation of protein complexes. On the interactions between viral and human proteins.
Pappin, D., Cold Spring Harbor Laboratory: MS 101: Fundamentals of mass spectrometry for proteomics.
Thompson, W., Duke Center for Genomic and Computational Biology, Durham, North Carolina: Label-free quant, AUC vs. spectral counting, tools for data interpretation.
Tomlinson, R., AstraZeneca, Waltham, Massachusetts: Introduction of inlysate chemoproteomics and work flow. Industrial case study.
Tsai, T-H., Northeastern University, Boston, Massachusetts: Statistics and experimental design for proteomic analyses.
Zhang, A., AstraZeneca, Waltham, Massachusetts: Introduction of inlysate chemoproteomics and work flow. Industrial case study.

Cellular Biology of Addiction

August 8–14

INSTRUCTORS **D. Belin,** University of Cambridge, United Kingdom
A. Bonci, National Institute on Drug Abuse, Baltimore, Maryland
C. Evans, University of California, Los Angeles
B. Kieffer, Douglas Hospital Research Centre, McGill University, Montreal, Quebec, Canada

ASSISTANTS **J. Bailly,** McGill University, Montreal, Quebec, Canada
M. Puaud, University of Cambridge, United Kingdom

Drug addiction is the most costly neuropsychiatric disorder faced by our nation. Acute and repeated exposure to drugs produces neuroadaptation and long-term memory of the experience, but the cellular and molecular processes involved are only partially understood. The primary objective of the proposed workshop was to provide an intense dialog on the fundamentals, state-of-the-art advances, and major gaps in the cell and molecular biology of drug addiction. Targeted to new or experienced investigators, the workshop combined formal presentations and informal discussions to convey the merits and excitement of cellular and molecular approaches to drug addiction research. With the advent of genomics and proteomics, an extraordinary opportunity now exists to develop comprehensive models of neuroadaptive processes fundamental to addiction, withdrawal, craving, relapse to drug use, and general brain function. A range of disciplines and topics were represented, including noninvasive brain imaging to identify drug targets and adaptive processes; neuroadaptive processes at the molecular and cellular level, neural networks and their modulation, the relevance of genotype to susceptibility and drug response; tolerance and adaptation at the cellular level; and approaches to exploiting the daunting volume of data generated by neuroinformatics. This workshop provided an integrated view of current and novel research on neuroadaptive responses to addiction,

fostered discussion on collaboration and integration, and provided critical information needed to construct a model of addiction as a disease and novel molecular targets for biological treatments. Beyond the plane of scientific endeavor, the information is vital for formulating public policy and for enlightening the public on the neurobiological consequences of drug use and addiction. This workshop was designed to generate interest in this level of analysis, open conduits for collaborations, and present novel routes to investigating the neurobiology of addictive drugs.

This course was supported with funds provided by the U.S. National Institute on Drug Abuse.

PARTICIPANTS

Bbosa, G., Ph.D., Makerere University College of Health Sciences, Kampala, Uganda
Bollati, F., Ph.D., Institute of Experimental Pharmacology IFEC-CONIC, Cordoba, Argentina
Caligiuri, S., Ph.D., Icahn School of Medicine at Mount Sinai Hospital, New York
Chioma, V., B.S., Medical University of South Carolina, Charleston
Conway, S., B.S., University of Illinois, Chicago
Gamble-George, J., Ph.D., University of Florida McKnight Brain Institute, Gainesville
Garcia Keller, C., Ph.D., Medical University of South Carolina, Charleston
Ghanbari, R., Ph.D., University of North Carolina, Chapel Hill
Hillmer, A., Ph.D., Yale University School of Medicine, New Haven, Connecticut
Justinussen, J., M.S., University of Copenhagen, Denmark
Lefevre, E., Ph.D., University of Minnesota, Minneapolis
Lichtenberg, N., B.S., University of California, Los Angeles
Linker, K., B.S., University of California, Irvine
Lobingier, B., Ph.D., University of California, San Francisco
Mendez, I., Ph.D., University of California, Los Angeles
Miszkiel, J., Ph.D., University of Texas Medical Branch, Galveston
Nennig, S., B.S., University of Georgia, Athens
O'Neill, P., Ph.D., Washington University School of Medicine in St. Louis, Missouri
Risher, M-L., Ph.D., Duke University, Durham, North Carolina
Williams, D., Ph.D., Johns Hopkins University, Baltimore, Maryland
Zakiniaeiz, Y., B.A., Yale University, New Haven, Connecticut
Zhang, P., Ph.D., University of Virginia, Charlottesville
Zhang, X., B.E., University of Mississippi Medical Center, Jackson

SEMINARS

Belin, D., University of Cambridge, United Kingdom: The psychological and neural basis of the vulnerability to addiction: Insights from preclinical studies.
Bonci, A., National Institute on Drug Abuse, Baltimore, Maryland: From synaptic plasticity and optogenetics to a novel treatment against cocaine abuse.
Bruchas, M., Washington University School of Medicine in St. Louis, Missouri: Dissecting neuromodulatory circuits and signaling in affective behavior.
Chavkin, C., University of Washington, Seattle: Molecular pharmacology of kappa opioids in addiction.
Evans, C., University of California, Los Angeles: Opioids and negative affect.
Golshani, P., University of California, Los Angeles, School of Medicine, Calabasas: New tools for imaging network dynamics in freely behaving animals.
Kalivas, P., Medical University of South Carolina, Charleston: Using the neurobiology of will power to cure addiction.
Kenny, P., Icahn School of Medicine at Mount Sinai, New York: Role for the habenula in nicotine seeking.
Kieffer, B., Douglas Hospital Research Centre, McGill University, Montreal, Quebec, Canada: Opioid receptors and brain function.
Kober, H., Yale University, New Haven, Connecticut: Human neuroscience of addiction: Focus on fMRI.
Nairn, A., Yale University School of Medicine, Guilford, Connecticut: Signal transduction processes in addiction.
Nestler, E., Mount Sinai University, New York: Transcriptional and epigenetic mechanisms of addiction.
Picciotto, M., Yale University, Guilford, Connecticut: Molecular mechanisms contributing nicotine addiction.
Stuber, G., University of North Carolina, Chapel Hill: Contemporary approaches for dissecting brain reward circuits.
Sunahara, R., University of California, San Diego, La Jolla: Manipulating G-protein-coupled receptors by drugs and hormones.
Witkowski, J., Cold Spring Harbor Laboratory: Ethics lecture.

Programming for Biology

October 16–31

INSTRUCTORS **S. Prochnik,** DOE/Joint Genome Institute, Walnut Creek, California
S. Robb, Stowers Institute for Medical Research, Kansas City, Missouri

ASSISTANTS **J. Bredeson,** University of California, Berkeley
M. Campbell, Cold Spring Harbor Laboratory
X. Doan, Stowers Medical Institute, Kansas City, Missouri
L. Kuderna, Institut de Biologia Evolutiva, Barcelona, Spain
D. Muna, Cold Spring Harbor Laboratory
D. Triant, University of Virginia, Charlottesville

More often than not, today's biologist is studying data that is too complex or numerous to be analyzed without a computer, and only boilerplate analysis can be performed with existing tools. Questions specific to the data set require novel analysis pipelines to be designed and written in computer code. Designed for lab biologists with little or no programming experience, this course gave students the bioinformatics and scripting skills necessary to exploit this abundance of biological data. The only prerequisite for the course was a strong commitment to learning basic UNIX and a scripting language.

This year, we offered the course in Python, an easy-to-learn scripting language with a growing code base and community of users. The course began with one week of introductory coding, continued with practical topics in bioinformatics—with plenty of coding examples—and ended with a group coding project. Formal instruction was provided on every topic by the instructors, teaching assistants, and invited experts. Students solved problem sets covering common scenarios

in the acquisition, validation, analysis, and visualization of biological data. They learned how to design, construct, and run powerful and extensible analysis pipelines in a straightforward manner. Final group projects were chosen from ideas proposed by students and were guided by faculty. Students were provided with a library of Python reference print and e-books that they were able to bring home with them.

The primary focus of this course was to provide students with practical programming experience, rather than to present a detailed description of the algorithms used in computational biology.

This course was supported with funds provided by the National Human Genome Research Institute. Access to cloud computational resources may be supported by an education grant from Amazon Web Services.

PARTICIPANTS

Barua, A., M.Sc., Okinawa Institute of Science and Technology, Japan
Bauchet, G., Ph.D., Boyce Thompson Institute, Ithaca, New York
Behrens, A., M.Sc., Max-Planck Institute for Biochemistry, Münich, Germany
Blanchard, A., Ph.D., University of Nottingham, Leicestershire, United Kingdom
Brewer, W., B.A., Duke University, Durham, North Carolina
Cenzer, M., Ph.D., University of North Carolina, Chapel Hill
Graf-Grachet, N., M.S., Oklahoma State University, Stillwater
Iwema, C., Ph.D., University of Pittsburgh, Pennsylvania
Kemmerer, Z., B.A., University of Wisconsin, Madison
Lavarias, M., B.S., University of Wisconsin, Madison
Miller, J., B.S., Medical College of Wisconsin, Milwaukee
Ngo, B., Ph.D., Weill Cornell/Memorial Sloan Kettering, New York
Singer, J., Ph.D., University of Utah, Salt Lake City
Sun, Y., B.S., Stanford and Carnegie Science for Plant Biology, California
Syed, S., Ph.D., University of Massachusetts Medical School, Worcester
Underbayev, C., Ph.D., National Institutes of Health, Bethesda, Maryland
Webb, S., B.A., University of Calgary, Alberta, Canada
Zhang, X., Ph.D., Columbia University, New York, New York

SEMINARS

Prochnik, S., Joint Genome Institute, Walnut Creek, California: Unix 1: Unix overview, the basics, advanced Unix, Unix cheat sheet. Python I: Python1 overview, running python, PyCharm, syntax, data types, and variables. Python III: Sequences, strings, lists, tuples. Python V: Iterables, I/O, and files. Python VII: Functions, scope, modules. Bioinformatics tools overview I: What's coming/Define bioinformatics concepts/Gene prediction, alignment, searching, and assembly. Python IX: Biopython.
Robb, S., Stowers Institute for Medical Research, Kansas City, Missouri: Unix 2: Text editors, Git for beginners. Python II: Operators, truth, logic, numbers. Python IV: Loops, dictionaries, sets. Python VI: Regular expressions. Python VIII: Exceptions, data structures.
Marques-Bonet, T., Institut Biologia Evolutiva, Barcelona, Spain: Structural variation.
Kuderna, L., Institut Biologia Evolutiva, Barcelona, Spain: Structural variation.
Cain, S., Ontario Institute for Cancer Research, Toronto, Canada: Bioinformatics tools overview II: GMOD and JBrowse.
Eilbeck, K., University of Utah, Salt Lake City: Ontologies: Controlled vocabulary.
Haas, B., Broad Institute, Northbridge, Massachusetts: RNA-Seq and transcript assembly.
Pearson, W., University of Virginia, Charlottesville: Sequence similarity I: Sequence similarity search I. Sequence similarity II: Sequence similarity search II/Alignment data problems.
Triant, D., University of Virginia, Charlottesville: Genome assembly.
Campbell, M., Cold Spring Harbor Laboratory: Next-gen sequencing: File formats/FASTA, FASTQ, BAM, VCF.
Witkowski, J., Cold Spring Harbor Laboratory: Ethics and reproducibility in research.

X-Ray Methods in Structural Biology

October 16–31

INSTRUCTORS
W. Furey, University of Pittsburgh, Pennsylvania
G. Gilliland, Janssen Research & Development, LLC, Spring House, Pennsylvania
A. McPherson, University of California, Irvine
A. Perrakis, Netherlands Cancer Institute, Amsterdam, Netherlands
J. Pflugrath, Rigaku Americas Corporation, The Woodlands, Texas

ASSISTANTS
R. Edayathumangalam, Weill Cornell Medicine, New York, New York
M. Whitley, University of Pittsburgh, Pennsylvania

X-ray crystallography has been the cornerstone of structural biology for half a century. This intense laboratory/computational course focused on the major techniques used to determine the three-dimensional structures of macromolecules. It was designed for scientists with a working knowledge of protein structure and function, but who were new to macromolecular crystallography. Topics that were covered included basic diffraction theory, crystallization (proteins, nucleic acids, complexes, and membrane proteins), synchrotron X-ray sources and optics, data collection and processing, structure solution by experimental phasing methods (SAD, MAD, MIR, and others) and molecular replacement, electron density maps improvement (solvent flattening, noncrystallographic averaging, etc.), model building and refinement, structure validation, coordinate deposition, and structure presentation. In addition, for the first time, the course extended to theory and computation for small angle X-ray scattering (SAXS) and single-particle cryoelectron microscopy.

Participants learned through extensive hands-on experiments in fully equipped labs, crystallized multiple proteins and determined their crystal structures by several methods while learning

through extensive lectures on theory. Informal discussions behind the techniques were frequent, and students were responsible also for collecting questions to be answered in specific sessions.

This course was supported with funds provided by the National Institute of General Medical Sciences.

PARTICIPANTS

Bakelar, J., Ph.D., Purdue University, West Lafayette, Indiana
Bantysh, O., Ph.D., Molecular Biology Institute of Barcelona, CSIC, Spain
Creixell, P., Ph.D., Massachusetts Institute of Technology, Cambridge
Cristobal, J., B.S., State University of New York, Buffalo
Harsini, F., B.Sc., Texas Tech University Health Sciences Center, Lubbock
Hinck, A., Ph.D., University of Pittsburgh, Pennsylvania
Johnson, J., B.A., University of Minnesota, Blaine
Lara-Tejero, M., Ph.D., Yale University, New Haven, Connecticut
Leeper, T., Ph.D., Kennesaw State University, Georgia
Lohry, D., B.S., University of Tennessee, Knoxville
Magala, P., Ph.D., University of Washington, Seattle
Minniberger, S., M.Sc., Leibniz-Institut für Molekulare Pharmakologie, Berlin, Germany
Rubinstein, R., Ph.D., Columbia University/HHMI, New York, New York
Sulon, S., B.S., Thomas Jefferson University, Philadelphia, Pennsylvania
Tseng, R., Ph.D., University of California, San Diego
Williamson, Z., B.S., University of Kentucky, Lexington

SEMINARS

McPherson, A., University of California, Irvine: Crystallization of macromolecules: Basic theory. Symmetry, periodicity, unit cells, space groups, Miller planes, and lattices. Waves, vectors, and complex numbers. Fundamental diffraction relationships and Bragg's law, diffraction patterns, reciprocal space, and Ewald's sphere. Fourier transforms and the electron density equation. Macromolecular crystals: Visualization of crystallization at the molecular level.

Newman, J., CSIRO, Parkville, Australia: Crystallization of macromolecules: Practical approaches. Tricks, tips, and modern approaches toward crystallization.

Caffrey, M., Trinity College, Dublin, Ireland: Membrane protein crystallization for structure-function studies using bicelles and lipid mesophases.

Gilliland, G., Janssen Research & Development, Spring House, Pennsylvania: Maximizing crystallization success through seeding.

Pflugrath, J., Rigaku Americas, The Woodlands, Texas: X-ray data collection and processing. Cryo-crystallography. Scaling and merging.

Sweet, R., Brookhaven National Laboratory, Shoreham, New York: A slightly different view of fundamental crystallography. X-ray sources and optics.

Holton, J., University of California, San Francisco: Beamline basics and tips and tricks for improving diffraction. More on Cryo. Data collection strategy.

Borek, D., University of Texas Southwestern Medical Center, Dallas: X-ray data processing. Data collection strategy. Advanced data processing. DIALS introduction.

Furey, W., University of Pittsburgh, Pennsylvania: Patterson space and Isomorphous replacement. Density modification: Solvent flattening, phase combination, NCS. Beamline introduction. Direct methods. Anomalous scattering, SAD and MAD.

Thorn A., University of Hamburg, Germany: An introduction to SHELXC/D/E. The basics of twinning in crystals of macromolecules.

Perrakis, A., Netherlands Cancer Institute, Amsterdam, Netherlands: Crystal mounting and data collection automation. Automated model building and rebuilding: From ARP/WARP. Automated model building and rebuilding: From PDB_REDO. Making presentation graphics technical overview…and assignments for own work to students. Macromolecular animation.

Whitley, M., University of Pittsburgh, Pennsylvania: XDS introduction.

Smith, C., Stanford University, California: Synchrotron data collection and femtosecond crystallography.

Witkowski, J., Cold Spring Harbor Laboratory: Ethics, reproducibility, and vigor.

Hammel, M., Lawrence Berkeley Laboratory, Berkeley, California: Small-angle scattering.

Terwilliger, T., Los Alamos National Laboratory, New Mexico: Automated structure solution and model building.

Read, R., University of Cambridge, United Kingdom: Molecular replacement: New structures from old. Using SAD data in Phaser.

Kleywegt, G., European Bioinformatics Institute, Cambridge, United Kingdom: Just because it's in Nature doesn't mean it's true… . The wonderful world of structure archiving!

Tronrud, D., The Light and Electron Works, Springfield, Oregon: Macromolecular refinement. Electron density maps.
Adams, P., Lawrence Berkeley Laboratory, Berkeley, California: Structure refinement. Low-resolution structure refinement.
Richardson, J., Duke University, Durham, North Carolina: Detection and repair of model errors using MolProbity. Presentation of structures history and perspectives.
Richardson, D., Duke University, Durham, North Carolina: Detection and repair of model errors using MolProbity.
Williams, C., Duke University, Durham, North Carolina: Detection and repair of model errors using MolProbity.
Conway, J., University of Pittsburgh, Pennsylvania: Cryo-electron microscopy introduction. Applications of cryo-electron microscopy.
Hendrikson, W., Columbia University, New York, New York: MAD and SAD phasing.
Emsley, P., MRC Laboratory of Molecular Biology, Cambridge, United Kingdom: Model-building tools in Coot.

Antibody Engineering, Phage Display, and Immune Repertoire Analysis

October 17–30

INSTRUCTOR **D. Siegel,** University of Pennsylvania, Philadelphia

CO-INSTRUCTOR **M.A. Pohl,** Tri-Institutional Therapeutics Discovery Institute, New York, New York

ASSISTANTS **S. Gilgunn,** National Institute of Bioprocess Research & Training, Dublin, Ireland
R. Güler, KTH Royal Institute of Technology, Stockholm, Sweden
D. Hernandez, Sackler Institute at NYU School of Medicine, New York

Advances in the generation and selection of antibodies from combinatorial libraries allow for the rapid production of antibodies from immune and nonimmune sources. This intensive laboratory/lecture course focused on the construction of combinatorial antibody libraries expressed on the surface of filamentous phage and the subsequent selection of desired antibodies from the library. Students learned the theoretical and practical aspects of constructing combinatorial libraries from immune and nonimmune sources as well as the construction of synthetic antibody libraries. Antigen-specific recombinant monoclonal antibodies were selected from the library. Production, purification, and characterization of antibody fragments expressed in *E. coli* were also covered.

The lecture series, presented by course faculty and a number of invited speakers, emphasized theory and practice of antibody display technologies, expression of antibodies in *E. coli* and mammalian cells, antibody structure and function, bacterial display of antibodies and other ligand-binding domains, the immunobiology of the antibody response, and the use of monoclonal antibodies for therapy—including the design of chimeric antigen receptor T cells. We also

discussed principles and protocols for generation and analysis of immune repertoires determined by next-generation sequencing.

This course was supported with funds provided by the National Institute of General Medical Sciences.

PARTICIPANTS

Bajic, G., Ph.D., Harvard Medical School, Boston, Massachusetts
Baron, S., B.S., University of Buffalo, New York
Bonfils, C., M.Sc., KisoJi Biotechnology, Laval, Quebec, Canada
De Leon, G., Ph.D., North Carolina State University, Raleigh
Hill, L., Ph.D., Adaptimmune Ltd, Oxford, United Kingdom
Hutchinson, M., Ph.D., National Institute on Aging, Baltimore, Maryland
Oldham, R., B.Sc., University of Toronto, Canada
Patsoukis, N., Ph.D., Beth Israel Deaconess Medical Center, Boston, Massachusetts
Rogovskyy, A., Ph.D., Texas A&M University, College Station
Schoenenwald, A., M.S., Medical University of Vienna, Austria
Schubert, R., M.D., University of California, San Francisco
Seo, H., Ph.D., Wayne State University, Detroit, Michigan
Szot-Karpinska, K., Ph.D., Institute of Physical Chemistry, PAS, Warsaw, Poland
Tano, H., M.S.E., AlbaNova Universitetscentrum, Stockholm, Sweden

SEMINARS

Boyd, S., Stanford University, California: Antibody repertoire analysis.
DeKosky, B., The University of Kansas, Lawrence: High-throughput analyses of paired antibody heavy- and light-chain repertoires.
Derda, R., University of Alberta, Edmonton, Canada: Genetically encoded chemically modified peptides.
Dreier, B., University of Zürich, Switzerland: In vitro evolution of proteins using ribosome display.
Lorenz, I., Tri-Institutional Therapeutics Discovery Institute, New York, New York: Antibody discovery at the Tri-I TDI.
Payne, A., University of Pennsylvania, Merion Station: Lineage tracing of isotype-specific B-cell repertoires in pemphigus.
Rader, C., The Scripps Research Institute, Jupiter, Florida: From phage display to cancer immunotherapy.
Silverman, G., New York University School of Medicine, New York: Development of the immune response and repertoire cloning. Characterization of dominant epitopes on staphylococcal toxins using phage display technology.
Witkowski, J., Cold Spring Harbor Laboratory: Ethics, reproducibility, and vigor.

Advanced Sequencing Technologies and Applications

November 7–18

INSTRUCTORS

M. Griffith, Washington University School of Medicine in St. Louis, Missouri
O. Griffith, Washington University School of Medicine in St. Louis, Missouri
E. Mardis, Nationwide Children's Hospital Research Institute, Columbus, Ohio
W. McCombie, Cold Spring Harbor Laboratory
A. Quinlan, University of Utah, Salt Lake City

ASSISTANTS

A. Farrell, University of Utah, Salt Lake City
S. Goodwin, Cold Spring Harbor Laboratory
J. Havrilla, University of Utah, Salt Lake City
M. Kramer, Cold Spring Harbor Laboratory
S. Kravitz, University of Utah, Salt Lake City
V. Magrini, Nationwide Children's Hospital, Columbus, Ohio
S. McGrath, Nationwide Children's Hospital, Columbus, Ohio
T. Sasani, University of Utah, Salt Lake City
A. Wagner, Washington University, Saint Louis, Missouri
J. Walker, Washington University School of Medicine in St. Louis, Missouri
R. Wappel, Cold Spring Harbor Laboratory
A. Ward, University of Utah, Salt Lake City

Over the last decade, massively parallel DNA sequencing has markedly impacted the practice of modern biology and is being utilized in the practice of medicine. The constant improvement of these platforms means that costs and data generation timelines have been reduced by orders of magnitude, enabling investigators to conceptualize and perform sequencing-based projects that heretofore were prohibitive in terms of time, cost, and required number of samples. Furthermore,

the application of these technologies to answer questions previously not experimentally approachable is broadening their impact and application. However, data analysis remains a complex and often vexing challenge, especially as data volumes increase.

This intensive two-week course explored the use and applications of massively parallel sequencing technologies, with a focus on data analysis and bioinformatics. Students were instructed in the detailed operation of several platforms, including library construction procedures, general data processing, and in-depth data analysis. A diverse range of the types of biological questions enabled by massively parallel sequencing technologies was explored, including DNA re-sequencing of known cancer genes, de novo DNA sequencing and assembly of genomes, RNA sequencing, and others that were tailored to the student's research areas of interest.

Cloud-based computing was also explored. Guest lecturers highlighted unique applications of these disruptive technologies. We encouraged applicants from a diversity of scientific backgrounds including molecular evolution, development, neuroscience, medicine, cancer, plant biology, and microbiology.

This course was supported with funds provided by the National Human Genome Research Institute.

Access to cloud computational resources may be supported by an AWS in Education Grant award from Amazon.

PARTICIPANTS

Bennett, C., M.D., Stanford University, Palo Alto, California
Caruso, L., Ph.D., Temple University, Philadelphia, Pennsylvania
Christensen, L., M.S., University of Southern Denmark, Odense M
Christie, E., Ph.D., Peter MacCallum Cancer Centre, Melbourne, Victoria, Australia
Cody, W., Ph.D., Texas A&M University, College Station
Dai, W., M.S., University of Massachusetts Medical School, Worcester
del Carmen Lafita Navarro, M., M.S./Ph.D., University of Texas Southwestern Medical Center, Dallas
de Miguel, F., Ph.D., Yale University, New Haven, Connecticut
Enogieru, O., Pharm.D., University of California, San Francisco
Gautreau, I., B.S., New England Biolabs, Ipswich, Massachusetts
Kaczmarek Michaels, K., Ph.D., Tufts University School of Medicine, Boston, Massachusetts
Lizama, C., M.S./Ph.D., University of California, San Francisco
Mancilla, V., Ph.D., University of North Texas Health Science Center, Fort Worth
Park, S., M.D., University of California, San Diego, La Jolla
Rosas, U., Ph.D., Universidad Nacional Autónoma de México, Mexico City
Senaratne, A., B.Sc., Institut Curie-Centre de Recherche, Paris, France
Silveira, P., B.S., Federal University of Rio de Janeiro, Brazil
Wojtas, A., B.S., Mayo Clinic, Jacksonville, Florida
Yan, J., Ph.D., Brigham and Women's Hospital, Boston, Massachusetts
Zhang, Z., B.Eng., Harvard University/Dana-Farber Cancer Institute, Boston, Massachusetts

SEMINARS

Mardis, E., Nationwide Children's Hospital, Columbus, Ohio: Overview of next-generation sequencing technologies. Cancer genomics and immunogenomics.
Witkowski, J., Cold Spring Harbor Laboratory: Ethics lecture.
Zody, M., New York Genome Center, New York: Computational biology at NYGC.
Maher, C., Washington University School of Medicine in Saint Louis, Missouri: Dissecting the RNA-interactome with NGS.
Preall, J., Cold Spring Harbor Laboratory: Single-cell sequencing technology and applications.
Schultz, N., Memorial Sloan Kettering Cancer Center, New York, New York: Identifying driver alterations and therapeutic options in cancer.
Meltz Steinberg, K., Washington University School of Medicine in St. Louis, Missouri: Introduction to NGS data analysis.
Griffith, O., Washington University School of Medicine in St. Louis, Missouri: Introduction to Cloud computing.

Marth, G., University of Utah, Salt Lake City: Variant discovery lecture. Tumor evolution and heterogeneity.
Quinlan, A. and Sasani, T., University of Utah, Salt Lake City: Disease variant discovery session with GEMINI. Genome arithmetic with BEDTOOLS and challenge problems.
Hoffman, M., Princess Margaret Cancer Centre, Toronto, Ontario, Canada: Data integration techniques.
Griffith, M., Washington University School of Medicine in St. Louis, Missouri: Intro to RNA-sequencing lecture.
Lappalainen, T., New York Genome Center & Columbia University, New York: Gene regulation, allelic expression, QC from GTeX.
Haas, B., Broad Institute, Northbridge, Massachusetts: Transcript assembly.
Sedlazeck, F., Johns Hopkins University, Baltimore, Maryland: Genome assembly with Illumina/PacBio/Nanopore.
Dewar, K., McGill University/Genome QC Innovation Centre, Montreal, Quebec, Canada: Intro to genome assembly of PacBio + Illumina data.
Nattestad, M., Cold Spring Harbor Laboratory: Genomic data visualization lecture.
Elemento, O., Weill Cornell Medicine, New York: Introduction to EpiGenomics analysis.
Goecks, J., Oregon Health and Science University, Portland: Analysis with Galaxy.

Scientific Writing Retreat

November 15–19

INSTRUCTORS **C. Lambert,** Cold Spring Harbor Laboratory
S. Matheson, Cell Reports, Cambridge, Massachusetts

WRITING COACHES **S. Gary,** Cold Spring Harbor Laboratory
J. Jansen, Cold Spring Harbor Laboratory
C. Martin, *Current Biology*, Cambridge, Massachusetts
J. Rubin, Columbia University, New York, New York
G. Teitzel, *Trends in Microbiology,* Cambridge, Massachusetts

The goal of this retreat was to have participants progress significantly on writing projects while improving their professional communication skills. It included a mix of formal sessions and less structured writing time. The formal sessions covered publication writing for scientific journals from the perspectives of Cell Press and Cold Spring Harbor Laboratory Press; writing clearly and conversationally about research in ways that engage diverse audiences—a skill particularly useful when developing lay summaries for NIH and NSF proposals; and style tips and considerations for clear professional writing in all forms.

The less structured sessions of the retreat included small writing groups and dedicated individual writing time. For the small group sessions, participants were preassigned to groups of three to four people for the purpose of soliciting peer feedback on writing samples they submitted ahead of time. For the individual writing sessions, coaches were on hand to work with participants one-on-one. As with all CSHL meetings and courses, participants were required to respect the confidentiality of any unpublished research they may have read during the retreat.

This course was supported with funds provided by the National Institutes of Health, National Institute of General Medical Sciences.

PARTICIPANTS

Acar, H., Ph.D., University of Chicago, Illinois
Baxter, A., Ph.D., Université de Montréal, Montreal, Quebec, Canada
Carrillo, R., Ph.D., The University of Chicago, Illinois
Chen, J., Ph.D., National Cancer Institute, Bethesda, Maryland
Dang, C., Ph.D., Columbia University, New York, New York
de Caceres Bustos, A., Ph.D., Duke University School of Medicine, Durham, North Carolina
Fujita, K., Ph.D., The Feinstein Institute for Medical Research, Manhasset, New York
George, J., Ph.D., Jackson Laboratory for Genomic Medicine, Farmington, Connecticut
Lee, J., Ph.D., University of Chicago, Illinois
Low, D., M.D., Johns Hopkins Medicine, Baltimore, Maryland
Luo, Y-J., Ph.D., Harvard University, Cambridge, Massachusetts
Rassouli-Taylor, L., Ph.D., Weill Cornell Medical College, New York, New York
Shetty, S., Ph.D., Rutgers University, Piscataway, New Jersey
So, N-Y., Ph.D., Columbia University/HHMI, New York, New York
Subhasappa, R.B., Ph.D., Iowa State University, Ames
van der Spoel, A., Ph.D., Dalhousie University, Halifax, Nova Scotia, Canada

SEMINARS

Lambert, C., Cold Spring Harbor Laboratory, and Matheson, S., Cell Press, Cambridge, Massachusetts: Session on top-10 tips. Lay summaries and writing for nonexpert audiences.
Martin, C., *Current Biology*, Cambridge, Massachusetts: The Cyrus Martin Show.
Matheson, S., Cell Press, Cambridge, Massachusetts: Publications and manuscripts.

Foundations of Comparative Genomics

November 29–December 6

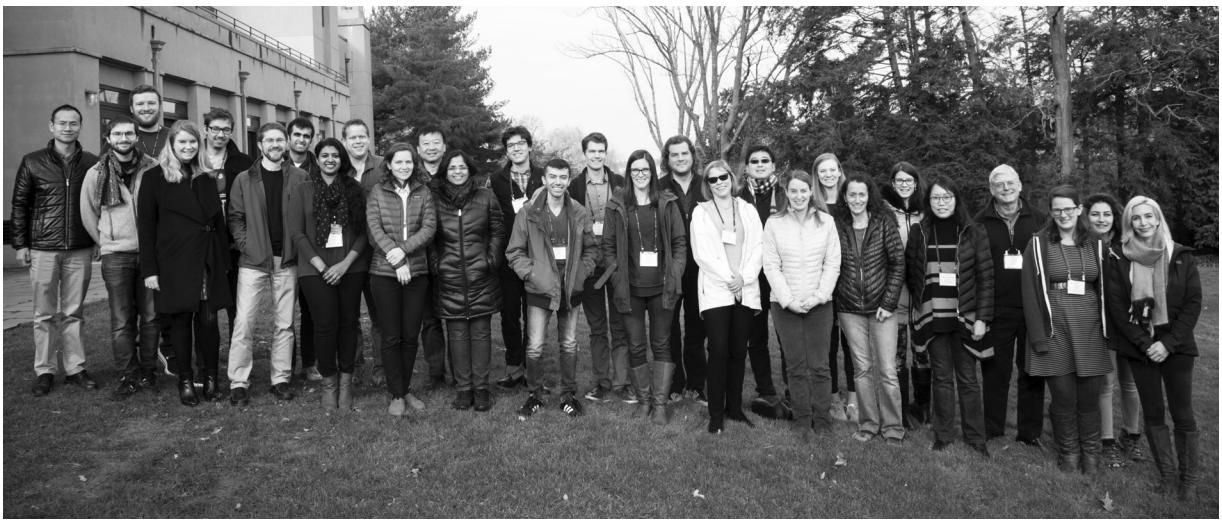
INSTRUCTORS
A. Mackey, University of Virginia, Morrisville
W. Pearson, University of Virginia, Charlottesville
L. Stubbs, University of Illinois, Urbana
J. Taylor, Johns Hopkins University, Baltimore, Maryland

ASSISTANTS
P. DeFord, Johns Hopkins University, Baltimore, Maryland
D. Miller, Stowers Institute for Medical Research, Kansas City, Missouri
O. Sabik, University of Virginia, Charlottesville
D. Triant, University of Virginia, Charlottesville

This course presented a comprehensive overview of the theory and practice of computational methods for the identification and characterization of functional elements from DNA sequence data. The course focused on approaches for extracting the maximum amount of information from protein and DNA sequence similarity through sequence database searches, statistical analysis, and multiple sequence alignment.

Additional topics included alignment and analysis of "next-gen" sequencing data, with applications from metagenomic, RNA-Seq, and CHiP-seq experiments; the Galaxy environment for high-throughput analysis; regulatory element and motif identification from conserved signals in aligned and unaligned sequences; integration of genetic and sequence information in biological databases; and genome browsers and genome features.

The course combined lectures with hands-on exercises; students were encouraged to pose challenging sequence analysis problems using their own data. It was designed for biologists seeking advanced training in biological sequence and genome analysis, computational biology core resource directors and staff, and individuals in other disciplines (e.g., computer science) who wished



to survey current research problems in biological sequence analysis. Advanced programming skills were not required. Its primary focus was the theory and practice of algorithms in computational biology, with the goals of using current methods more effectively for biological discovery and developing new algorithms.

This course was supported with funds provided by the National Human Genome Research Institute.

PARTICIPANTS

Alavattam, K., B.S., Cincinnati Children's Hospital Medical Center, Ohio
Alvarez, J., Ph.D., New York University, New York
Arat, S., Ph.D., The Jackson Laboratory Genomic Medicine, New Britain, Connecticut
Chen, M., Ph.D., University of California, Riverside
Danielne Sandor, K., Ph.D., Sanford Burnham Prebys Medical Discovery Institute, Orlando, Florida
Ferguson, A., M.S., University of Pittsburgh, Pennsylvania
FitzGerald, E., Ph.D., Brown University, Providence, Rhode Island
Gillani, R., M.D., Boston Children's Hospital, Boston, Massachusetts
Glover, H., B.Sc., University of Sydney, Camperdown, New South Wales, Australia
Iyer, S., M.Sc., Stony Brook University/CSHL, New York
Kerwin, R., Ph.D., Michigan State University, East Lansing
Lee, H.H-C., M.Sc., Princess Margaret Hospital, Laichikok, Hong Kong
Luo, H., Ph.D., University of Florida, Gainesville
Mitzelfelt, K., Ph.D., University of Washington, Seattle
Podolsky, M., M.D., University of California, San Francisco
Rodríguez-Carballo, E., Ph.D., University of Geneva, Switzerland
Rothstein, M., B.S., Cornell University, Ithaca, New York
Shetty, S., Ph.D., Rutgers University, Piscataway, New Jersey
Wolf, A., B.S., Cedars-Sinai Medical Center, Los Angeles, California
Yuan, S., B.S., The Rockefeller University, New York, New York

SEMINARS

Pearson, W., University of Virginia, Charlottesville: Protein evolution and sequence similarity searching. Practical sequence similarity searching. PSSMs, HMMs, and phenotype prediction. Multiple sequence alignment.
Mackey, A., HemoShear, LLC, Charlottesville, Virginia: Genome annotation (HMM basics). Gene lists to pathways. RStudio visualization.
Taylor, J., Johns Hopkins University, Baltimore, Maryland: Sequencing technologies: New genomics. Assembling genomes and transcriptomes. Galaxy for high-throughput analysis. Variation and SNP discovery. Probing higher-dimension chromatin structure.
Stubbs, L., University of Illinois, Urbana: Regulatory genomics.
Leek, J., Johns Hopkins University, Baltimore, Maryland: RNA sequencing technologies. Statistics for genomes I.
Miller, D., Stowers Institute for Medical Research, Kansas City, Missouri: Genome assembly and analysis using nanopore.
Hawkins, D., University of Washington, Seattle: Chromatin states 1: Analysis of histone modifications. Chromatin states 2: Overlapping data sets.
Wilson Sayres, M., Arizona State University, Tempe: Sex bias in reference-based alignments.
Mahony, S., Pennsylvania State University, University Park: Genomics of gene regulation 1: Analyzing protein–DNA-binding interactions. Genomics of gene regulation 2: Characterizing transcription-factor-binding dynamics.

The Genome Access Course

INSTRUCTORS **D. Fagegaltier,** New York Genome Center, New York
A. Gordon, New York Genome Center, New York
E. Hodges, Vanderbilt University School of Medicine, Nashville, Tennessee
B. King, University of Maine, Orono, Maine
S. Munger, The Jackson Laboratory, Bar Harbor, Maine

The Genome Access Course was an intensive two-day introduction to bioinformatics. The course was broken into modules that are each designed to give a broad overview of a given topic, with ample time for examples chosen by the instructors. Each module featured a brief lecture describing the theory, methods, and tools, followed by a set of worked examples that students completed. Students were encouraged to engage instructors during the course with specific tasks or problems that pertained to their own research. The core of the course was the analysis of sequence information framed in the context of completed genome sequences. Featured resources and examples primarily come from mammalian species, but concepts can be applied to any species. The course also featured methods to assist the analysis and prioritization of gene lists from large-scale gene expression and proteomics experiments. Summaries of the topics are listed below.

The topic Sequence, Gene, and Protein Resources included NCBI sequence, gene, and protein resources; model organism databases—mouse genome informatics, rat genome database, ZFIN, FlyBase; protein sequence and domain resources—UniProt, PDB, InterPro; proteomics resources—IPI, ExPaSY, PRIDE; microRNA resources—miRBase, microCosm Targets, TargetScan, PicTa; repositories of high-throughput sequence data; repositories of gene expression data: GEO, ArrayExpress; gene expression profiling resources; and gene ontology.

The topic Genome Browsers included genome sequencing and assembly; gene annotation; overview and comparison of major genome browsers—Ensembl, UCSC, NCBI Map Viewer; adding custom tracks; and bulk genome retrieval tools—BioMart, UCSC table browser.

The topic De Novo Analysis of Sequences included local, global, pairwise, and multiple sequence alignments; BLAST and BLAT algorithms; scoring matrices—PAM, BLOSUM; iterative

profile and pattern searches; multiple sequence alignment programs; and visualizing and editing multiple alignments.

The topic Sequence Variation included types of sequence and structural variation; SNP resources: dbSNP; and structural variation resources: dbVar, DGVa, HGVbase.

The topic Comparative Genome Analysis and Functional Genomic Elements included finding putative regulatory elements in genome sequence by comparing genomes; ortholog and paralog resources; Multicontigview in Ensembl; comparative tracks in the UCSC genome browser; and DCODE and ENCODE resources.

The topic Analysis of High-Throughput Sequence Data included common file formats—FASTQ, SAM, BAM; quality control and diagnostic analyses; mapping reads to a reference sequence; finding putative mutations and polymorphisms; RNA-Seq data analysis; ChIP-seq data analysis; de novo assembly; and Galaxy resources.

The topic Gene Set Enrichment and Pathway Analysis included prioritizing genes from microarray and proteomics experiments; gene set enrichment analysis tools—GSEA, DAVID; pathway resources—Reactome, HPRD NetPath, KEGG; and protein interaction resources—MIPS, MINT, BIND, DIP.

Each student was provided with a laptop (if needed) and internet access for the duration of the course. Students were also allowed to bring their own laptop to the course provided it met the following requirements: (1) a standard browser (Internet Explorer, Firefox, etc.) that is up-to-date with security patches and bug fixes, (2) wireless internet capacity, and (3) the ability to view and modify plain text files and spreadsheets (e.g., Microsoft Word and Excel). Both PCs and Macs were acceptable as long as they were updated with all security patches and bug fixes.

The curriculum of The Genome Access Course has been developed in conjunction with staff at the Wellcome Trust Sanger Institute and the European Bioinformatics Institute (Hinxton, United Kingdom), who teach a parallel series of courses in the United Kingdom. For more information, see the Open Door Workshops.

This course was supported with funds provided by the Helmsley Charitable Trust.

March 13–15

Students: 31

PARTICIPANTS

Appert, E., Children's Hospital of Philadelphia, Pennsylvania
Chaligne, R., Weill Cornell Medicine, New York, New York
Chang, H.R., New York University Langone Medical Center, New York
Chen, Q., Lieber Institute for Brain Development, Baltimore, Maryland
Docherty, A., University of Utah School of Medicine, Salt Lake City
Doran, E., Genentech, South San Francisco, California
ElGamal, D., Ohio State University, Columbus
Fan, X., Saint Louis University, Missouri
Hu, X., University of Pennsylvania, Philadelphia
Hudson, A., Yale University School of Medicine, New Haven, Connecticut
Jain, P., Children's Hospital of Philadelphia, Pennsylvania
John, J., Stony Brook University Hospital, New York
Kim, S.Y., New York University Langone Medical Center, New York
Korb, E., The Rockefeller University, New York, New York
Kutlu, M., Pennsylvania State University, University Park
Layman, D., Estee Lauder Companies Inc., Melville, New York
Lee, S., Abilene Christian University, Texas
Li, D., University of Pennsylvania, Philadelphia
Lin, W-H., Columbia University, New York, New York
MacFarland, S., Children's Hospital of Philadelphia, Pennsylvania
Manners, M., University of Pennsylvania, Philadelphia
Nkhisang, T., Harvard T.H. Chan School of Public Health, Boston, Massachusetts
Patino, G., Oakland University William Beaumont School Medicine, Rochester, Michigan
Punzi, G., Lieber Institute for Brain Development, Baltimore, Maryland
Rizvi, N., Merck & Co., Kenilworth, New Jersey
Rubi, T., University of Michigan, Ann Arbor

Shaw, A., Yale University School of Medicine, New Haven, Connecticut
Sonawane, P., Children's Hospital of Philadelphia, Pennsylvania
Tang, K.H., New York University, New York
Ursini, G., Lieber Institute for Brain Development, Baltimore, Maryland
Wilson, J., Tufts Medical Center, Boston, Massachusetts

SEMINAR

Schneider, V., National Center for Biotechnology Information, Bethesda, Maryland: The NCBI genome data viewer. Reference genome assembly.

November 13–15

Students: 34

PARTICIPANTS

Adams, T., Cold Spring Harbor Laboratory
Afelik, S., University of Illinois, Chicago
Anyanwu, A., University of Michigan Medical School, Plymouth Meeting, Pennsylvania
Bendriem, R., Weill Cornell Medicine, New York, New York
Bhatnagar, P., SRI International, Menlo Park, California
Chen, M., University of California, Riverside
Cheon, H., Cleveland Clinic, Cleveland, Ohio
Datta, M., Massachusetts General Hospital, Boston
Dunn, M., Cold Spring Harbor Laboratory
Furtado, R., Albert Einstein College of Medicine, Bronx, New York
Gilchrist, A., Midwestern University, Downers Grove, Illinois
Grigorenko, E., Baylor College of Medicine, Houston, Texas
Hansen, L., Emory University, Atlanta, Georgia
Irmady, K., The Rockefeller University, New York, New York
Lipsky, A., Weill Cornell Medicine, New York, New York
Maddox, S., McLean Hospital/Harvard Medical School, Belmont, Massachusetts
Menon, M., Icahn School of Medicine at Mount Sinai, Irvington, New York
Nicholson, H., Dana-Farber Cancer Institute, Boston, Massachusetts
Obeng, E., St. Jude Children's Research Hospital, Memphis, Tennessee
Olsen, M., Virginia Tech, Blacksburg
Padhi, B., Health Canada, Ottawa, Ontario, Canada
Reese, J., University of Colorado, Aurora
Shetty, S., Rutgers University, Piscataway, New Jersey
Shor, E., Rutgers University, Newark, New Jersey
Simone-Finstrom, M., USDA-ARS, Baton Rouge, Louisiana
Stevenson, J., University of Pittsburgh, Pennsylvania
Storch, G., Washington University School of Medicine in St. Louis, Missouri
Sun, Z., Yale University School of Medicine, New Haven, Connecticut
Tarui, T., Tufts University School of Medicine, Boston, Massachusetts

Taylor, J., Memorial Sloan Kettering Cancer Center, New York, New York
Tuua, R., Missouri State Public Health Laboratory, Jefferson City
Walsh, S., Rollins College, Winter Park, Florida
Wang, J.X., Cold Spring Harbor Laboratory
Zhao, D., MD Anderson Cancer Center, Houston, Texas

SEMINAR

Fagegaltier, D., New York Genome Center, New York: The role of noncoding RNAs in regulating gene expression during neurodegenerative disease.

The Laboratory acknowledges the generosity of the following companies that loaned equipment and reagents to the various courses:

3i
10× Genomics
AB SCIEX LLC
ADInstruments Inc.
Advanced Analytical
Advanced Illumination
Agilent Technologies Inc.
ALA Scientific Instruments
A-M Systems
Andor Technologies Inc.
Applied Scientific Instrumentation Inc.
BD Biosciences
Berthold Technologies USA, LLC
BioLegend
Bioline USA
Bio-Rad Laboratories
BioTek
Biotool
Bitplane
BMG Labtech
Bruker Daltonics Inc.
Cairn Research Ltd.
Charles River Laboratories Inc.
Chroma Technology Corporation
CNRS
Coherent
ConOptics
Covaris, Inc.
CrystaLaser
Diagenode
Drummond Scientific Company
eBioscience
EMD Millipore Corp.
Electron Microscopy Sciences
Epicentre Technologies Corp.
Eppendorf North America
Excelitas Technologies Corp.
Exiqon, Inc.
Fotodyne Inc.
GE Healthcare
Hamamatsu Photonic Systems
Hamilton Drayage, Inc.
Hampton Research
Harlan
Harvard Apparatus Inc.
Heka Instruments Inc.
HOLOEYE
Homebrew
Illumina Inc.
Intan Technologies
Intelligent Imaging Inc.
Invitrogen
Keyence
Kinetic Systems Inc.
Labcyte
Leica Biosystems
Leica Microsystems Inc.
LI-COR
Life Technologies
Lumen Dynamics
Lumencor
MDS Analytical Technologies
MiTeGen
Molecular Devices Corp.
Morrell Instrument Co.
Nanodrop Technologies
NanoTemper Technologies
Narishige International USA Inc.
Nasco
New England Biolabs Inc.
Newport Corporation
Nikon Inc.
NSK America Corporation
OKO-Lab
PerkinElmer Life and Analytical Sciences
Photometrics
Physik Instruments
Precisionary Instruments
Prior Scientific
Promega Corporation
QIAGEN Inc.
Quantitative Imaging Corp.
RC Testing Services
Sage Science Inc.
Scientifica Ltd.
Singer Instruments
Sutter Instruments
Taconic Farms Inc.
Tecan Group US, Inc.
The Jackson Laboratories
Thermo Fisher Scientific
Thor Labs
Tokai Hit
TotalLab Ltd.
Waters Corporation
White Labs
World Precision Instruments
Zenith Biotech

SEMINARS

INVITED SPEAKERS PROGRAM ("CSHL SEMINAR SERIES")

Each year, Cold Spring Harbor Laboratory invites speakers from outside the institution to present their findings. These weekly seminars keep the CSHL staff current on the latest scientific developments and broaden their perspectives. Graduate students and postdoctoral fellows meet with the seminar speakers for lunch immediately after the seminar, providing an opportunity for the exchange of ideas in an informal setting.

Speaker	Title	Host
January		
Hidde Ploegh, Ph.D., Professor, Department of Microbiology and Immunobiology, Harvard Medical School/ Research Associate, Department of Medicine, Division of Molecular Biology, Boston Children's Hospital	Using nanobodies: Phenotypic screens, visualization of an immune response	Leemor Joshua-Tor
Paula Arlotta, Ph.D., Professor, Harvard University	Development of the cerebral cortex: From the embryo to 3D brain organoids	Camila dos Santos
Leslie Vosshall, Ph.D., Robin Chemers Neustein Professor and Howard Hughes Medical Institute Investigator, The Rockefeller University	Neurobiology of the world's most dangerous animal (McClintock Lecture)	CSHL WiSE
February		
Joseph Ecker, Ph.D., Professor, Salk International Council, Chair in Genetics, Howard Hughes Medical Institute, The Salk Institute for Biological Studies	Epigenomic diversity and the regulatory DNA landscape	Doreen Ware
Hopi Hoekstra, Ph.D., Alexander Agassiz Professor of Zoology, Howard Hughes Medical Institute Investigator, Harvard University	Evo and devo: How the mouse got its stripes	CSHL Graduate Students
Yang Dan, Ph.D., Assistant Professor, Department of Molecular and Cell Biology, University of California, Berkeley	Neural circuits controlling sleep	Josh Huang
Jennifer Lippincott-Schwartz, Ph.D., Group Leader, Howard Hughes Medical Institute/Janelia Research Campus	Emerging fluorescence technology to study the spatial and temporal dynamics of organelles within cells	Richard Sever
March		
John Ioannidis, M.D., D.Sc., Professor of Medicine, Health Research and Policy, Biomedical Data Science and Statistics, Stanford University	What does reproducible basic and preclinical research mean?	Jesse Gillis
Karen Adelman, Ph.D., Professor, Harvard Medical School	Making sense of noncoding transcription	Camila dos Santos
Carol Greider, Ph.D., Daniel Nathans Professor and Director, Department of Molecular Biology and Genetics, Johns Hopkins University School of Medicine/Bloomberg Distinguished Professor, Department of Biology, Johns Hopkins University	Telomeres and telomerase: Setting the equilibrium (McClintock Lecture)	CSHL WiSE

Speaker	Title	Host
Eva Nogales, Ph.D., Howard Hughes Medical Institute Investigator & Professor, University of California, Berkeley	Molecular visualization of the transcription initiation process by cryo-EM	Leemor Joshua-Tor
Joanna K. Wysocka, Ph.D., Professor, Department of Chemical and Systems Biology, Howard Hughes Medical Institute Investigator, Stanford University School of Medicine	Gene regulatory mechanisms in human development and evolution	Terri Grodzicker
April		
Dominique Bergmann, Ph.D., Professor of Biology, Howard Hughes Medical Institute Investigator, Stanford University	Making a difference: Optimizing pattern formation in the plant epidermis	Zach Lippman
October		
Leif Ellisen, M.D., Ph.D., Professor at Harvard Medical School, Massachusetts General Hospital Cancer Center	Progenitor landscapes controlling epithelial carcinogenesis	Robert Maki
Yigong Shi, Ph.D., Professor, Vice Chancellor, Tsinghua University, Beijing, China	Mechanistic insights into pre-mRNA splicing by spliceosome	Adrian Krainer and Leemor Joshua-Tor
November		
Amir Amedi, Ph.D., Associate Professor, Department of Medical Neurobiology, Hebrew University of Jerusalem and the Edmond and Lily Safra Center for Brain Sciences	The plastic topographic human brain cortex: From seeing with the ears to creating novel sensory experiences	Graduate Students
Barak Cohen, Ph.D., Professor, Department of Genetics, Washington University, St. Louis	Integration of local and regional *cis*-regulatory information in the genome	Adam Siepel
Jef Boeke, Ph.D., Director, Institute for System Genetics, NYU Langone Medical Center	Writing genomes	Molly Hammell
December		
Richard Ebright, Ph.D., Board of Governors, Professor of Chemistry and Chemical Biology, Rutgers University	RNA polymerase: The molecular machine of transcription	Leemor Joshua-Tor
Catherine Dulac, Ph.D., Higgins Professor of Molecular and Cellular Biology, Harvard University, Howard Hughes Medical Institute Investigator	Neurobiology of social behavior circuits	Pavel Osten
Anthony Fitzpatrick, Ph.D., Assistant Professor, Department of Biochemistry and Molecular Biophysics, Mortimer B. Zuckerman Mind Brain Behavior Institute, Columbia University	Cryo-EM structures of Tau filaments from Alzheimer's disease brain: Implications for fibril propagation arising from patient-based structural biology	Cat Donaldson

IN-HOUSE SEMINAR PROGRAM

Cold Spring Harbor Laboratory's In-House Seminars were initiated to provide a semiformal avenue for communication among the various research groups at the Laboratory. The seminars also afford a necessary opportunity for graduate students and postgraduate staff to develop their skills in organizing, presenting, and defending their research.

Speaker	Title
January	
Ivan Iossifov	Genetics of autism
February	
Anirban Paul	Unity in diversity: Transcriptional architecture of neuronal cell types
March	
Dawid Nowak	Drivers of metastasis in prostate cancer
Bo Li	The amygdala controls fear, anxiety, and beyond
Zach Lippman	Tomato: Secrets revealed
Rafaella Sordella	What doesn't kill you will make you stronger
April	
Pavel Osten	Neuroanatomy in the 21st century: Can we do better than Ramón y Cajal?
Jesse Gillis	Transcriptional identity from cell to self
October	
Partha Mitra	Tipping points in network performance: Phase transitions in machine learning and distributed control
David M. McCandlish	Modeling complex sequence–function relationships
November	
Andrea Schorn	LTR-retrotransposon control by tRNA-derived small RNAs
Doug Fearon	CXCR4-mediated T-cell exclusion: A pathway to "immune privilege" in cancer
December	
Chris Vakoc	Transcriptional vulnerabilities of the cancer cell
Anne Churchland	Excitatory and inhibitory neural populations reflect single trial decisions
Mikala Egeblad	Turning to the dark side: Immune responses promoting cancer

BANBURY CENTER

BANBURY CENTER EXECUTIVE DIRECTOR'S REPORT

The year 2017 was one of transition for the Banbury Center: It was with exhilaration that I took the baton from Jan Witkowski, who spent 30 remarkable years cultivating critical scientific discourse at the Center. Despite the change in leadership, activities at Banbury continued to be guided by the Center's mission to further scientific knowledge and the well-being of society. The year also marked 40 years since the Banbury Center was officially opened. In his speech, "How Scientists Work," at the 1977 dedication ceremony, Francis Crick pointed to small meetings as the best way for scientists to share and inspire new ideas and strategies (McElheny 2003). Despite technological advances in communication that have allowed a more interconnected world, Crick's sentiment still proves true with each small meeting convened at Banbury. The breadth of Banbury meetings in 2017, spanning discovery and translational science, public health, policy, education, and innovation, reflected the ever-growing need for multisector and multidisciplinary engagement at small meetings across a broad range of issues.

By the Numbers

In 2017, the Banbury Center hosted 46 activities and events, including Banbury meetings, laboratory retreats, and courses directed by the Watson School of Biological Sciences (Immunology; Physical Biology of the Cell) and the Meetings and Courses Program (Workshop on Pancreatic Cancer; Vision: A Platform for Linking Circuits, Behavior, and Perception; Neural Data Science; Autism Spectrum Disorder; Cellular Biology of Addiction; Workshop on Leadership in Bioscience; Scientific Writing Retreat).

A total of 536 individuals took part in Banbury meetings, with 72% marking their first occasion. Participants were drawn from 28 countries, spread across six continents; Antarctica continues to elude representation at Banbury meetings. The 79% of participants from the United States spanned 36 states and territories. The largest portion of Banbury meeting attendees, representing academic organizations (71%), participants from industry (15%), other not-for-profit organizations (9%), governments (4%), and publishing/writing (1%), brought diverse perspectives and new cross-sector relationships. Banbury continues to strive for gender diversity: Women represented 31% of 2017 participants, and we aim to improve this ratio in the coming years.

In 2017 Banbury continued to attract financial support from across sectors, with the largest funding drawn from not-for-profit organizations (59%). For-profit organizations constituted 30% of funding, with the Cold Spring Harbor Laboratory Corporate Sponsor Program (CSP) accounting for more than half of that figure. The CSP funds were absolutely vital to ensuring that Banbury was able to convene cutting-edge meetings in 2017; we continue to be grateful to those member organizations and to Cat Donaldson in recruiting membership and Michelle Corbeaux for coordinating participation in meetings.

Neuropharmacology and Human Stem Cell Models, September

Discovery and Translational Science

Banbury's year began with *Chemiexcitation in Human Disease and Aging*, a meeting that exemplified multidisciplinary engagement, with experts from chemistry, pathology, aging, and neurodegeneration among those brought together to explore the mechanisms and principles underlying chemiexcitation's pathological consequences. In the same way, *NLRs Sans Frontières* brought the plant and animal research communities together to exchange new and unpublished data on a microbial recognition mechanism common to both kingdoms, the NLR proteins. Although the *Enhanceropathies: Enhancer Function Variation in Animal Development, Morphological Variation, and Disease* meeting convened a relatively less diverse group of experts, it likewise sought to identify core principles underlying biological function (or dysfunction), focusing in this case on consequences of changes to noncoding regulatory sequences in DNA.

Two meetings targeted regulated cell death, kicking off in April with *Ferroptosis: A Critical Review* and circling back after Thanksgiving for *Regulated Necrosis: Pathways and Mechanisms*. The former centered on a relatively new addition to the necrotic cell death family, so-named because of its dependence on iron. Highly productive discussions from the meeting shaped a review paper (see publications list at the end of this report), and the burgeoning field of researchers has now outgrown Banbury; a Cold Spring Harbor Asia meeting will pick up the conversation in November 2018. The second meeting, held in autumn, targeted the most well characterized mode of regulated cell death: necroptosis. This pathway is implicated in viral infections as well as cancer, ischemic injury, and a number of inflammatory conditions. Participants at this meeting were challenged to share their newest research findings to inspire new ideas to move the field forward.

Cancer was in the crosshairs in 2017, with two meetings asking new questions in established research areas. In April, *Better Cancer Therapy from Redox Biology* assessed the complexity of redox regulation in the context of cancer biochemistry and therapy. Later in the year, potential mechanisms underlying anticancer effects of an established drug were explored in *Metformin: Translating Biology into the Clinic*.

Three meetings built on Banbury's strong history in neuroscience. September's *Neuropharmacology and Human Stem Cell Models* meeting centered on the latest advances in generating human stem cell–derived disease models and resources for therapeutic discovery. The ALS Association's meeting in October—*Cell Biology of ALS: Emerging Themes from Human Genetics*—identified

new strategies in the fight against this neurological disease, focusing on new research to uncover detailed pathways involved in vesicular trafficking, autophagy, DNA damage, and neuroinflammation. Finally, the year's meetings concluded with *Post-Traumatic Neuroinflammation: Roles in Pathogenesis of Long-Term Consequences and Repair*, which convened researchers studying the cellular and molecular responses to injury and infection, with physician scientists treating patients affected by stroke, traumatic brain injury, and perinatal brain injury to share data and identify areas for collaboration.

Public Health

Among the earliest Banbury Center meetings were those confronting issues of environmental and public health, including the hazards of cigarette smoking (1979) and concerted efforts in combating AIDS (1983). These are areas in which Banbury-style meetings are especially critical to convening multidisciplinary groups, from experts in pathogenicity and toxicity to epidemiologists, social scientists, community health workers, and policy makers. Banbury's 2017 schedule included three meetings in this vein. May's *Maximizing Impact of New HIV Prevention Technologies in Sub-Saharan Africa,* supported by the Bill & Melinda Gates Foundation, framed a productive strategy session around three broad questions: What do we know?, What do we need to know?, and What do we need to do next? The group's diversity in sector, field, research, and in-country experience was critical, especially during the final sessions in which participants developed a model to integrate major conclusions and recommendations. Key findings were presented at the World STD and HIV Congress.

Although the HIV meeting focused on prevention strategies heavily based on risk and barriers to uptake, two additional meetings paired public health needs with new scientific strategies. The first centered on Lyme disease, well timed in a year with predictions of escalating outbreaks. *Protective Immunity and Vaccines for Lyme Disease,* made possible through the support of the Steven and Alexandra Cohen Foundation, explored best strategies for vaccine development, with lively discussion despite a forceful autumn storm and subsequent power outage across the estate. Capping Lyme disease activities in 2017, a report on next-generation Lyme disease diagnostics was published—an output of a 2016 meeting (see publications list for detail). The second meeting, July's *Opportunities for Reduction of Aflatoxin Contamination of Food*, tackled the fungal contamination of food supplies, a global issue affecting health, agriculture, and trade. Experts at this

M. Levine, B. Ren, A. Rada-Iglesias

Banbury meeting examined opportunities to reduce aflatoxin contamination through production of abiotic stress-resistant crop varieties, biocontrol, and host-induced gene silencing.

Policy, Strategy, and Education

The pastoral setting of the Center and tradition of inspiring discussions in a confidential environment make Banbury an ideal site for important policy discussions, internal strategy development, and intensive training. In addition to the summer courses, we once again welcomed Boehringer Ingelheim Fonds for their North American retreat, *Communicating Science*, during which the foundation's Ph.D. fellows spent nearly 1 week learning and applying tools across the communication spectrum. Two other returning groups were the Integrated Translational Science Center (ITSC) and the Lustgarten Foundation, convening members and external experts to monitor progress, develop strategy, and strengthen collaboration. Funded by the National Cancer Institute, the ITSC works to bridge the gap between bench and bedside, bringing together innovative research at Cold Spring Harbor Laboratory and the Jackson Laboratory with the clinical expertise of SWOG's physician investigators. Representatives of the three groups convened at Banbury to discuss pressing challenges in oncology and to brainstorm collaborative projects to address unmet needs. Similarly working toward a world without cancer, the Lustgarten Foundation returned to the Conference Room for their 2017 Scientific Meeting, which provided an opportunity for the Scientific Advisory Board, as well as Foundation-supported investigators, to discuss research and strategy, identifying the most promising new avenues to bolster progress in the field. Marking their first Banbury visit, *Project Santa Fe* convened leadership from innovative clinical laboratories in March to develop strategies for services that maximize impact, improve patient outcomes, and cut costs.

S. Schutzer, A. Marques, J.W.R. Hovius

The year also found Banbury embracing the surge of innovation and entrepreneurship in the biosciences: Banbury joined efforts with the Keystone for Incubating Innovation in Life Sciences Network (KIILN) for July's *Foundation2017*. Tents were erected on the estate to host nearly 70 bioscience entrepreneurs, investors, translational researchers, and industry leaders for a 1-day retreat that included panel sessions and plenty of informal networking. This nontraditional event was an excellent opportunity for attendees from the tristate area to experience Banbury while discussing practical challenges in bio-entrepreneurship and broadening their networks.

It Takes a Village

Finally, it is with great humility that I acknowledge those who keep the Banbury Center running at such a high level, and who were critical to ensuring continuity of quality during my transition to director. Michelle Corbeaux and Pat Iannotti power the meetings and events with expert coordination and organization. Basia Polakowski oversees our three residence buildings, ensuring our guests are comfortable, while the Culinary Services team keeps them well fed, and the Audiovisual staff ensures technology supports rather than distracts. Jose Pena Corvera, John Shea, and Paulo Krizanovski look after 55 acres of impeccable grounds, and the entire Facilities team quite literally keep us running. Hakon Heimer has continued to be essential to the development of Banbury's pipeline of mental illness–focused meetings, and our extensive collaboration with the Meetings and Courses Program broadens Banbury's portfolio of activities.

Of course, the Banbury Center's mere existence and international reputation are owed to Charles Robertson and the Robertson family, to Bruce Stillman and James Watson, and especially to my esteemed predecessor, Jan Witkowski.

Rebecca Leshan
Executive Director

PUBLICATIONS

Branda JA, Body BA, Boyle J, Branson BM, Dattwyler RJ, Fikrig E, Gerald NJ, Gomes-Solecki M, Kintrup M, Ledizet M, Levin AE, et al. 2017. Advances in serodiagnostic testing for Lyme disease are at hand. *Clin Infect Dis* doi: 10.1093/cid/cix943.

Jakubowski H. 2017. Homocysteine editing, thioester chemistry, coenzyme a, and the origin of coded peptide synthesis. *Life (Basel)* **7:** E6.

Korf BR, Blitzer MG, Demmer LA, Feldman GR, Watson MS. 2017. Report on the Banbury Summit Meeting on medical genetics training in the genomics era [Commentary]. *Genet Med* **19:** doi: 10.1038/gim.2017.38.

McElhenu VK. 2003. *Watson and DNA: Making a scientific revolution*, p. 170. Perseus Publishing, Cambridge, MA.

NordForsk 2017. *Nordic biobanks and registers: A basis for innovative research on health and welfare.* Policy paper, ISSN 1504–8640. Oslo, Norway.

Sanacora G, Heimer H, Hartman D, Mathew SJ, Frye M, Nemeroff C, Robinson Beale R. 2017. Balancing the promise and risks of ketamine treatment for mood disorders. *Neuropsychopharmacology* **42:** 1179–1181.

Stockwell BR, Friedmann Angeli JP, Bayir H, Bush AI, Conrad M, Dixon SJ, Fulda S, Gascón S, Hatzios SK, Kagan VE, et al. 2017. Ferroptosis: A regulated cell death nexus linking metabolism, redox biology and disease. *Cell* **171:** 273–285.

BANBURY CENTER MEETINGS

Dates	*Title*	*Organizer(s)*
February 12–15	Chemiexcitation in Human Disease and Aging	E. Bechara, D. Brash
February 24–March 1	BIF Seminar "Communicating Science"	C. Walther
March 19–22	Enhanceropathies: Enhancer Function Variation in Animal Development, Morphological Variation, and Disease	M. Levine, J. Wysocka
March 28–31	Project Santa Fe: 2nd Annual Meeting	Project Santa Fe Executive Committee
April 2–5	Ferroptosis: A Critical Review	X. Jiang, B.R. Stockwell
April 10–13	Better Cancer Therapy from Redox Biology	C. Chio, D. Tuveson
May 16–19	Maximizing Impact of New HIV Prevention Technologies in Sub-Saharan Africa	D. Pillay, H. Ward
June 12–14	Integrated Translational Science Center Workshop	L. Baker, L. Ellis, E. Liu, A. Schott, D. Tuveson
July 7	KiiLN Foundation 2017: Bio-Entrepeneurship in NYC	D. Brand, N. McKnight
July 9–12	Opportunities for Reduction of Aflatoxin Contamination of Food	J. Harvey, R. Michelmore, R. Nelson
September 10–13	Neuropharmacology and Human Stem Cell Models	N. Brandon, Z. Cader, S. Haggarty
September 17–20	NLRs Sans Frontières	J. Dangl, J. Jones, R. Vance
September 24–27	Metformin: Translating Biology into the Clinic	N. Chandel, V. Stambolic
October 22–24	Cell Biology of ALS: Emerging Themes from Human Genetics	L. Bruijn, A. Gitler, E. Holzbaur
October 29–November 1	Protective Immunity and Vaccines for Lyme Disease	E. Fikrig, S. Schutzer
November 12–14	Lustgarten Foundation Scientific Advisory Board Meeting	D. Tuveson, R. Vizza, A. Whiteley
November 26–29	Regulated Necrosis: Pathways and Mechanisms	D. Green, A. Linkermann
December 6–8	Post-Traumatic Neuroinflammation: Roles in Pathogenesis of Long-Term Consequences and Repair	R. Ransohoff, A. Schaefer, D. Schafer, J. Witkowski

BANBURY CENTER MEETINGS

Chemiexcitation in Human Disease and Aging

February 12–15

FUNDED BY **The LEO Foundation of Ballerup, Denmark, with additional support from the São Paulo Research Foundation (FAPESP) and L'Oréal**

ARRANGED BY **E. Bechara,** University of São Paulo and Federal University of São Paulo, Brazil
D. Brash, Yale University, New Haven, Connecticut

Chemiexcitation, the chemical excitation of electrons, is the biophysical process underlying bioluminescence. It had not been observed in mammals until a recent report demonstrated that chemiexcitation sends melanocytes down the path to melanoma. The same chemistry can occur in any tissue that contains melanin, suggesting that chemiexcitation may be an unknown step in neurodegenerative disorders such as Parkinson's disease, Alzheimer's disease, macular degeneration, and noise- or drug-induced deafness. The challenges are to identify, for each tissue, the sources of reactive nitrogen and oxygen species that initiate chemiexcitation; the melanin chemistry that creates the excited state; the DNA or protein alterations caused by energy transfer from, or chemical reaction with, the excited molecule; and the contribution of these alterations to pathogenic events. This multidisciplinary meeting brought together experts to build candidate pathways for each disease and a list of principles underlying the fields, suggesting experimental tools to address this previously unrecognized mode of disease.

Welcoming Remarks: **J.A. Witkowski,** Banbury Center, Cold Spring Harbor Laboratory

T. Sarna, H. Sies, E. Gaillard, E. Bastos

SESSION 1: Overview: A New Mode of Pathogenesis

Chairperson: D. Brash, Yale School of Medicine, New Haven, Connecticut

D. Brash, Yale School of Medicine, New Haven, Connecticut: Excited electrons in melanoma and beyond.

E. Bechara, University of São Paulo and Federal University of São Paulo, Brazil: Short history of chemiexcitation: Liaisons between dioxygen, carbonyls, and light.

SESSION 2: Excited States and Chemiexcitation

Chairperson: J-S. Taylor, Washington University, St. Louis, Missouri

B. Kohler, Ohio State University, Columbus: Excited states from light.

L. Blancafort, University of Girona, Spain: Cyclobutane dimers and energy wormholes.

W. Joseph Baader, University of São Paulo, Brazil: Excited states from chemical and biochemical reactions.

SESSION 3: Before the Excited State: The Ingredients

Chairperson: K. George, L'Oréal, Clark, New Jersey

H. Sies, Heinrich Heine University, Dusseldorf, Germany: The redox code and adventures of reactive oxygen and nitrogen.

T. Sarna, Jagiellonian University, Krakow, Poland, and L. Zecca, Italian National Research Council, Milan, Italy: Melanin, neuromelanin, and their reactions.

SESSION 4: After the Excited State: The Weapon

Chairperson: G. Timmins, University of New Mexico, Albuquerque

E. Bastos, University of São Paulo, Brazil: Excited-state reactions: Where the energy goes.

T. Sarna, Jagiellonian University, Krakow, Poland: Singlet oxygen.

G. Wondrak, University of Arizona, Tucson: Ground-state carbonyls are still reactive: Glycation damage and excited states.

SESSION 5: Diseases of Tissues Containing Melanocytes

Chairperson: T. Sarna, Jagiellonian University, Krakow, Poland

J. Shupp, Medstar Washington Hospital Center, Georgetown University, Washington, D.C.: Wound healing and hypertrophic scars.

J. O'Malley, Massachusetts General Hospital, Boston: The cochlea and noise/drug-induced deafness.

E. Gaillard, Northern Illinois University, DeKalb: Retinal pigment epithelium and macular degeneration.

SESSION 6: Diseases of Tissues Containing Neuromelanin

Chairperson: J. Costa, Yale School of Medicine, New Haven, Connecticut

W. Surewicz, Case Western Reserve University, Cleveland, Ohio: Biophysics of prions and amyloid.

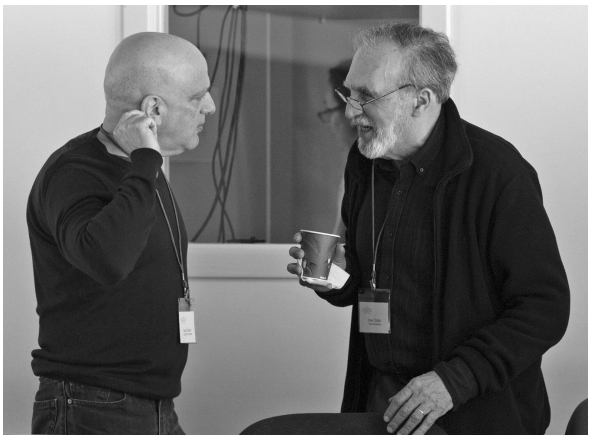
D. Sulzer, Columbia University, New York, and L. Zecca, Italian National Research Council: Parkinson's disease and progressive supranuclear palsy.

A. Vortmeyer, Indiana University, Indianapolis: Alzheimer's disease and late-age Down's syndrome.

SESSION 7: Assembling Pathways Breakout Groups

Chairpersons: D. Brash, Yale School of Medicine, New Haven, Connecticut, **K. George,** L'Oréal Advanced Research, Clark,

D. Brash



D. Sulzer, J. Costa

New Jersey, and **J. Costa,** Yale School of Medicine, New Haven, Connecticut

1. Deafness and Scars
2. Macular Degeneration
3. Parkinson's Disease and Progressive Supranuclear Palsy
4. Alzheimer's Disease and Late-Age Down's Syndrome

SESSION 8: Blocking the Pathways

Chairperson: B. Kohler, Ohio State University, Columbus

H. Sies, Heinrich Heine University, Dusseldorf, Germany: Antioxidant strategies: Enzymes and bioactives.

W. Joseph Baader and E. Bastos, University of Sao Paulo, Brazil: Intercepting and quenching electronically excited states.

L. Blancafort, University of Girona, Spain: Deactivating triplet states rapidly.

SESSION 9: The Précis

Chairperson: D. Sulzer, Columbia University, New York

Participants identified principles, methods, unanswered questions, and future experiments for the following:

1. Excited State Chemistry
2. ROS/RNS Chemistry, Reactive Carbonyls, Antioxidants
3. Pathology

BIF Seminar "Communicating Science"

February 24–March 1

FUNDED BY **Boehringer Ingelheim Fonds**

ARRANGED BY **C. Walther,** Boehringer Ingelheim Fonds, Mainz, Germany

The Boehringer Ingelheim Fonds (BIF) has an international fellowship program supporting outstanding Ph.D. students. Among the opportunities provided to fellows is rigorous training in communication through an annual retreat. It was a great pleasure to have them return in 2017 for interactive instruction in matters such as oral presentations and writing papers. This year's retreat marked the tenth such visit to Banbury.

Opening Remarks: C. Walther, Boehringer Ingelheim Fonds, Mainz, Germany

K. Achenback, Boehringer Ingelheim Fonds, Mainz, Germany: Communication: Why and how?

A. Katsnelson, Freelance Biomedical Writer and Editor, Northampton, Massachusetts: Writing techniques and how to structure papers.

W. Tansey, Vanderbilt University, Nashville, Tennessee: Preparing and delivering a scientific talk.

PowerPoint Presentations, Videotaped with Replay, and Feedback

K. Grace, Weill Cornell Medicine, New York: Image beautification and the slippery slope to misconduct.

M. Krzywinski, British Columbia Cancer Agency, Vancouver, British Columbia: Design of scientific concept and data figures with Adobe Illustrator.

C. Walther, Boehringer Ingelheim Fonds, Mainz, Germany: All about BIF.

Enhanceropathies: Enhancer Function Variation in Animal Development, Morphological Variation, and Disease

March 19–22

FUNDED BY **Cold Spring Harbor Laboratory Corporate Sponsor Program**

ARRANGED BY **M. Levine,** Princeton University, New Jersey
J. Wysocka, Stanford School of Medicine, California

Changes to noncoding regulatory DNA sequences can lead to transcriptional variation that, in turn, mediates inter- and intraspecies phenotypic divergence. Genetic alterations in enhancer sequences, or those remodeling their chromosomal context, can result in human disease including congenital malformations, cancer, and neurodegenerative and autoimmune disorders. This meeting convened experts who compared diverse processes and systems to reveal unifying principles underlying functional enhancer variation, such as optimization or loss of activator elements, modulation of repressive inputs, and alterations in long-range enhancer-promoter communication.

Welcoming Remarks: R. Leshan, Banbury Center, Cold Spring Harbor Laboratory

Overview of Meeting Objectives: M. Levine, Princeton University, New Jersey, and **J. Wysocka,** Stanford School of Medicine, California

SESSION I: Enhancer Variation in Development and Evolution

Chairperson: M. Levine, Princeton University, New Jersey

E. Farley, University of California, San Diego: Regulatory principles governing enhancer specificity.

D. Stern, Janelia Research Campus, Ashburn, Virginia: Evolution of transcription through a deep dive into the functional evolution of the shavenbaby enhancers.

E. Furlong, European Molecular Biology Laboratory, Heidelberg, Germany: Functional insights into chromatin topology and gene expression during embryonic development.

C. Danko, Cornell University, Ithaca, New York: Natural selection has shaped coding and noncoding transcription in primate $CD4^+$ T cells.

J. Wysocka, M. Levine, F. Spitz, J. Crocker

A. Stark, C. Rushlow

SESSION II: General Mechanisms of Enhancer Function

Chairperson: C. Rushlow, New York University

J. Wysocka, Stanford School of Medicine, California: On the dangers of mistaking correlation for causation: Histone modifications in enhancer function.

A. Stark, Research Institute of Molecular Pathology, Vienna, Austria: Decoding transcriptional regulation.

M. Levine, Princeton University, New Jersey: Enhancer–enhancer interactions within complex genes.

K. Adelman, Harvard Medical School, Boston, Massachusetts: Making sense of nonsense: A roadmap for deciphering the potential functions of noncoding RNAs.

R. Young, Whitehead Institute for Biomedical Research, Cambridge, Massachusetts: Things we don't understand about enhancers.

SESSION III: Transcription Factors

Chairperson: A. Stark, Research Institute of Molecular Pathology, Vienna, Austria

H. Bussemaker, Columbia University, New York: Accurate and sensitive quantification of protein–DNA recognition.

C. Rushlow, New York University: Zelda-binding sites as quantitative timers of target gene transcription.

J. Taipale, Karolinska Institutet, Stockholm, Sweden: Genome-wide analysis of protein–DNA interactions.

SESSION IV: Topology and Nuclear Architecture

Chairperson: E. Furlong, European Molecular Biology Laboratory, Heidelberg, Germany

F. Spitz, Institut Pasteur, Paris, France: Managing functional interactions between distant enhancers and genes.

B. Ren, Ludwig Institute for Cancer Research, La Jolla, California: An essential role of CTCF in chromatin organization?

A. Rada-Iglesias, Center for Molecular Medicine, Cologne, Germany: Polycomb proteins as topological facilitators of enhancer regulatory function.

G. Blobel, Children's Hospital of Philadelphia, Pennsylvania: Chromatin readers and nuclear architecture.

SESSION V: Enhancer Malfunction in Cancer

Chairperson: K. Adelman, Harvard Medical School, Boston, Massachusetts

B. Bernstein, Massachusetts General Hospital and Harvard, Charlestown: Hypermethylation and insulator dysfunction in cancer.

A. Shilatifard, Northwestern University, Chicago, Illinois: Enhancer biology and enhanceropathies in cancer.

P. Scacheri, Case Western University, Cleveland, Ohio: Aberrant enhancer activation in cancer progression.

SESSION VII: Wrap-Up and Next Steps

Meeting Conclusions

Project Santa Fe: 2nd Annual Meeting

March 28–31

FUNDED BY **Northwell Health**

ARRANGED BY **Project Santa Fe Executive Committee**

Project Santa Fe is a coalition of leadership from innovative clinical laboratories: Northwell Health Laboratories, Geisinger Health System, Henry Ford Health System, Kaiser Permanente Northern California Health Systems, and TriCore Reference Laboratories. We were pleased to host this group at Banbury in March to develop strategies for clinical lab testing services that maximize impact, improve patient outcomes, and cut costs. The outputs of this meeting were subsequently presented at two major conferences.

Geisinger Health System (Danville, Pennsylvania)
C. Christenson, M.D.
J. Olson, M.D.
S. Snyder, Ph.D.
M. Wilkerson, M.D.
D. Wolke, M.D.

Henry Ford Health System (Detroit, Michigan)
I. Rubinfeld, M.D.
G. Sharma, M.D.
J.M. Tuthill, M.D.
J. Waugh, M.D.
R. Zarbo, M.D.

Northwell Health (Lake Success, New York)
D. Breining, M.D.
T. Chang, M.D.
J. Crawford, M.D., Ph.D.
Y. Jacobs (observer)
T. Kothari, M.D., M.P.H.
L. Lomsadze
R. Miller
A. Murray
S. Roychoudhury, M.D. (observer)
R. Stallone
J. Yim

TriCore Laboratories (Albuquerque, New Mexico)
M. Crossey, M.D.
M. Dodd, PharmD
N. Fisher
K. Shotorbani

Special Participants
R. Michel, President and CEO, The Dark Report
M. Trusheim (Moderator), MIT Sloan School of Management; President, Co-Bio Consulting, LLC

Ferroptosis: A Critical Review

April 2–5

FUNDED BY **Burroughs Wellcome Fund, Cold Spring Harbor Laboratory Corporate Sponsor Program, Collaborative Medicinal Development, Memorial Sloan Kettering Cancer Center, and Ono Pharmaceutical Co. Ltd.**

ARRANGED BY **X. Jiang,** Memorial Sloan Kettering Cancer Center, New York
B.R. Stockwell, Columbia University, New York

Ferroptosis is a form of regulated, nonapoptotic cell death that involves overwhelming lipid peroxidation. Originally reported in 2012, ferroptosis has been of increasing interest because of its integration with cellular metabolism and its suggested role in cell death associated with degenerative diseases, carcinogenesis, stroke, intracerebral hemorrhage, traumatic brain injury, ischemia-reperfusion injury, and kidney degeneration in mammals, as well as heat stress in plants. This Banbury meeting brought together, for the first time, leading researchers working on diverse aspects of ferroptosis to explore mechanisms underlying this emerging form of regulated cell death and to suggest tools and guidelines for future studies.

Welcoming Remarks: R. Leshan, Banbury Center, Cold Spring Harbor Laboratory

Overview of Meeting Objectives: B. Stockwell, Columbia University, New York, and **X. Jiang,** Memorial Sloan Kettering Cancer Center, New York

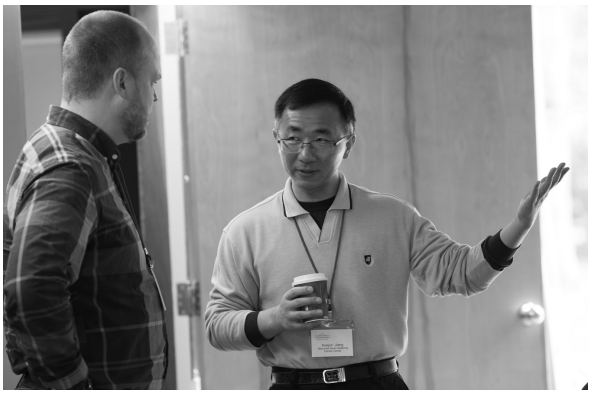
SESSION I: Regulators of Ferroptosis

Chairperson: B. Stockwell, Columbia University, New York

B. Stockwell, Columbia University, New York: Overview and update on ferroptosis.

M. Conrad, Helmholtz Zentrum München, Germany: In vivo mechanisms of ferroptosis control by GPX4.

D. Tang, University of Pittsburgh, Pennsylvania: P53 limits ferroptosis by blocking DPP4 activity.



S. Dixon, X. Jiang

M. Murphy, C. Prives, C. Rosenfeld

SESSION II: Ferroptosis and Metabolism

Chairpersons: X. Jiang, Memorial Sloan Kettering Cancer Center, New York, and **K. Salnikow,** NCI, National Institutes of Health, Bethesda, Maryland

X. Jiang, Memorial Sloan Kettering Cancer Center, New York: Metabolism and ferroptosis.

S. Dixon, Stanford University, California: The regulation of polyunsaturated fatty acid oxidation during ferroptosis.

SESSION III: ROS in Ferroptosis

Chairperson: V.E. Kagan, University of Pittsburgh, Pennsylvania

V.E. Kagan, University of Pittsburgh, Pennsylvania: ROS in ferroptosis.

K. Woerpel, New York University: Development of cyclic peroxides that induce ferroptosis, not apoptosis.

S. Hatzios, Yale University, New Haven, Connecticut: Oxidative cROSstalk at the host–microbe interface: A possible role for ferroptosis in infectious disease.

SESSION IV: Ferroptosis and Degenerative Disease: Part 1

Chairperson: A. Linkermann, Universitätsklinikum Carl Gustav Carus, Germany

A. Linkermann, Universitätsklinikum Carl Gustav Carus, Germany: The in vivo relevance of ferroptosis.

Q. Ran, University of Texas Health Science Center at San Antonio: GPX4 and neurogeneration.

SESSION V: Ferroptosis and Degenerative Diseases: Part 2

Chairperson: A Bush, University of Melbourne, Australia

A. Bush, University of Melbourne, Australia: The role of ferroptosis in Alzheimer's disease.

H. Bayir, University of Pittsburgh, Pennsylvania: What is the role of ferroptosis in acute brain injuries?

J. Pedro Friedmann Angeli, Helmholtz Zentrum München, Germany; Role of Acsl4 as a pro-ferroptotic gene.

SESSION VI: Ferroptosis and Cancer

Chairperson: M. Murphy, the Wistar Institute, Pennsylvania

M. Murphy, the Wistar Institute, Pennsylvania: An African-specific variant of p53 is defective for ferroptosis.

C. Prives, Columbia University, New York: Role of the p53 network in the regulation of ferroptosis.

S. Toyokuni, Nagoya University Graduate School of Medicine, Japan: Role of iron in carcinogenesis and tumor biology.

SESSION VII: Ferroptosis and Other Cell Fates

Chairperson: S. Fulda, Goethe Universitat, Germany

S. Fulda, Goethe Universitat, Germany: RSL3 and Erastin differentially regulate redox signaling to promote Smac mimetic-induced cell death.

S. Gascón, Ludwig-Maximilians University of Munich, Germany: ROS and ferroptosis in cell reprogramming.

B. Stockwell, S. Fulda

SESSION VIII: Ferroptosis in Diverse Contexts

Chairperson: M Overholtzer, Memorial Sloan Kettering Cancer Center, New York

G. Pagnussat, IIB-CONICET-National University of Mar del Plata, Argentina: A conserved cell death pathway across kingdoms: Ferroptosis in plants.

M. Overholtzer, Memorial Sloan Kettering Cancer Center, New York: Nanoparticle-mediated ferroptosis induction in cancer.

SESSION IX: Back to the Starting Point: Iron and Ferroptosis

Chairperson: D. Zhang, University of Arizona, Tucson

F. Torti, University of Connecticut, Storrs: Iron metabolism.

S. Torti, University of Connecticut, Storrs: The "fer" in ferroptosis.

D. Zhang, University of Arizona, Tucson: NRF2: An integrator of cellular iron and redox signaling.

SESSION X: Wrap-Up and Next Steps

Chairpersons: B. Stockwell, Columbia University, New York, and **X. Jiang**, Memorial Sloan Kettering Cancer Center, New York

Better Cancer Therapy from Redox Biology

April 10–13

FUNDED BY **The Oliver Grace Chair Fund**

ARRANGED BY **C. Chio,** Cold Spring Harbor Laboratory
D. Tuveson, Cold Spring Harbor Laboratory

An unanswered question in human health is whether anti-oxidation prevents or promotes cancer. Anti-oxidation has historically been viewed as chemopreventive, but emerging evidence suggests that antioxidants may be supportive of neoplasia. Leveraging cellular redox changes toward the development of a safe and effective therapeutic strategy necessitates experimental delineation of specific redox signaling pathways that are uniquely required by cancer cells to grow and to survive. This Banbury meeting focused on the complexity of redox regulation in the context of cancer biochemistry and therapy, exploring ROS genesis and metabolism in cancer cells, as well as the "productive" and "destructive" signal transduction by free radicals through the oxidation of intermediates.

Welcoming Remarks: R. Leshan, Banbury Center, Cold Spring Harbor Laboratory

SESSION I: Free Radicals and Antioxidants in Physiological Functions

Chairperson: A. Holmgren, Karolinska Institute, Stockholm, Sweden

J. Watson, Cold Spring Harbor Laboratory: To overcome chemoresistant cancers, use natural product quinones.

C. Winterbourn, University of Otago, Christchurch, New Zealand: Cellular mechanisms for regulating hydrogen peroxide metabolism and oxidative stress.

U. Jakob, University of Michigan, Ann Arbor: Role of polyphosphate in oxidative stress defense.

N. Tonks, Cold Spring Harbor Laboratory: Redox regulation of protein tyrosine phosphatases for therapeutic development.

C. Winterbourn, T. Dick

T. Mak, I. Chio

N. Chandel, Northwestern University, Chicago, Illinois: Functional genomic screens to uncover redox biology.
P. Schumacker, Northwestern University, Chicago, Illinois: Mitochondrial regulation of cell proliferation.

SESSION II: Free Radicals and Antioxidants in Cancer

Chairpersons: A. Ostman, Karolinska Institute, Stockholm, Sweden, and **K. Liby,** Michigan State University, East Lansing

K. Vousden, Francis Crick Institute, London, United Kingdom: Modulating TIGAR to probe ROS functions in tumor development and metastasis.
T. Mak, University of Toronto, Ontario, Canada: Modulation of oxidative stress as an anticancer strategy.
E. Schmidt, Montana State University, Bozeman: Endogenous oxidants and cellular antioxidant systems in liver cancer.
M. Bergo, Karolinska Institutet, Huddinge, Sweden: Antioxidants cause long-term programming of lung cancer cells into a metastatic phenotype.
N. Hay, University of Illinois, Chicago: Akt, hexokinase 2, ROS, and cancer therapy.

SESSION III: NRF2 in Redox Homeostasis and Metabolism

Chairperson: M. Espey, National Cancer Institute, Rockville, Maryland

J. Hayes, University of Dundee, United Kingdom: The mechanisms of repression of transcription factor Nrf2 and its cross talk with lipid metabolism.
C. Chio, Cold Spring Harbor Laboratory: Nrf2 promotes mRNA translation in pancreatic cancer.
M. Yamamoto, Tohoku University, Sendai, Japan: Molecular basis of Keap1-Nrf2 system and cancer.
T. Papagiannokopolus, New York University Medical School: Pro-tumorigenic NRF2 antioxidant program causes defects in central carbon metabolism.
G. DeNicola, Moffitt Cancer Center, Tampa, Florida: Compartmentalization of ROS production and metabolism.
D. Zhang, University of Arizona, Tucson: NRF2: An integrator of cellular iron and redox signaling.

SESSION IV: Redox Imaging

Chairperson: T. Dick, German Cancer Research Center, Heidelberg, Germany

V. Belousov, Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry: Metabolic engineering tools and fluorescent probes for redox biology.
K. Brindle, University of Cambridge, United Kingdom: Imaging oxidative stress in vivo.
C. Chang, University of California, Berkeley: Chemical imaging and proteomics probes for studying redox biology.
T. Dick, German Cancer Research Center, Heidelberg, Germany: Understanding the "anti-oxidant" *N*-acetyl cysteine.
Y. Yang, East China University of Science and Technology, Shanghai: Genetically encoded sensors for redox biology and their applications in drug screening.
M. Murphy, MRC Mitochondrial Biology Unit Cambridge, United Kingdom: Therapeutic alteration to the mitochondrial redox environment.

SESSION V: Therapeutics

Chairperson: T. Miller, IC-MedTech, Las Vegas, Nevada

D. Boothman, University of Texas Southwestern Medical Center, Dallas: Leveraging NQO1 bioactivatable drugs for tumor-selective ROS production and antitumor activity.
G. Buettner, University of Iowa, Iowa City: Using science to guide clinical trials for cancer treatment where redox biology is at the center.
D. Spitz, University of Iowa, Iowa City: $O_2^{\bullet-}$ and H_2O_2-mediated disruption of Fe metabolism causes the differential susceptibility of NSCLC and GBM cancer cells to pharmacological ascorbate.

S. Morrison, University of Texas Southwestern Medical Center, Dallas: Distant metastasis requires cancer cells to adapt to cope with oxidative stress.

E. Parkinson, University of Illinois, Urbana: Deoxynyboquinones as NQO1-targeted anticancer compounds.

SESSION VI: Wrap-Up and Next Steps

Chairpersons: D. Tuveson and J. Watson, Cold Spring Harbor Laboratory

K. Vousden, K. Brindle, U. Jakob

Maximizing Impact of New HIV Prevention Technologies in Sub-Saharan Africa

May 16–19

FUNDED BY **Bill & Melinda Gates Foundation**

ARRANGED BY **D. Pillay,** Africa Health Research Institute, Durban, South Africa
H. Ward, Imperial College London, United Kingdom

Despite the promotion of HIV combination prevention encompassing structural, behavioral, and biomedical interventions, HIV incidence in adolescent girls and young men and women in sub-Saharan Africa remains high. Recent data suggest that "treatment as prevention" approaches alone will be inadequate to limit the epidemic in this setting. A wide-ranging group of participants at this Banbury meeting were challenged to examine current knowledge about who is at risk, what drives that risk, and what facilitates or obstructs the uptake of preventive interventions—and to articulate what we need to know to get new and old preventive technologies taken up and used. During the meeting's final sessions, the group applied their discussions toward specific examples of high-risk individuals and concluded with the development of a model to integrate conclusions and recommendations.

Welcoming Remarks: R. Leshan, Banbury Center, Cold Spring Harbor Laboratory

Overview of Meeting Objectives: D. Pillay, Africa Health Research Institute, Durban, South Africa, and **H. Ward,** Imperial College London, United Kingdom

SESSION I: Introductions and Framework for the Meeting

Chairpersons: H. Ward, Imperial College London, United Kingdom, and **D. Pillay,** Africa Health Research Institute, Durban, South Africa

E. Emini and G. Garnett, Bill & Melinda Gates Foundation, Seattle, Washington: The Gates Foundation HIV Prevention Strategy.

H. Ward, Imperial College London, United Kingdom, and D. Pillay, Africa Health Research Institute, Durban, South Africa: Overview of the population (adolescent girls and young men and women in sub-Saharan Africa) and the technologies (new and old) that we will be considering.

R. Barnabas, University of Washington, Seattle: Incidence of HIV in adolescent girls, young women, and young men in sub-Saharan Africa: Who are the highest risk populations, 2016 UNAIDS data?

SESSION II: Who Is at Risk and Why?

Chairperson: H. Ward, Imperial College London, United Kingdom

J. Wamoyi, National Institute for Medical Research, Mwanza, Tanzania: Transactional sex and HIV among AGYW in sub-Saharan Africa.

S. Mojola, University of Colorado, Boulder: How social and ecological contexts shape and complicate individual decision-making about HIV prevention.

SESSION III: User Perspectives on HIV Prevention Technologies

Chairperson: M. Shahmanesh, University College London and Africa Health Research Institute, United Kingdom

J. Seeley, London School of Hygiene & Tropical Medicine, United Kingdom: Why men do and don't use male condoms.

M. Warren, AVAC, New York: Why women do and don't use female condoms.

M. Gafos, MRC, University College London, United Kingdom: Lessons from microbicide research: User perspectives, acceptability, and adherence.

H. Ward, Imperial College London, United Kingdom: Obstacles and facilitators: Summary of evidence on existing technologies.

SESSION IV: Understanding the Market for HIV Prevention Technologies: Products and Population Segmentation

Chairperson: H. McDowell, ViiV Healthcare, Brentford, United Kingdom

A. Gomez, AVAC, New York: Products, platforms, and people: Developing and delivering prevention that works.

K. Hallman, The Population Council, New York: Segmentation of sexual partner types: Results of participatory research.

SESSION V: Understanding HIV Prevention Behaviors: Theoretical Insights

Chairperson: G. Dallabetta, Bill & Melinda Gates Foundation, Washington, D.C.

R. Prasad, Final Mile Consulting, Chicago, Illinois: Using a behavioral economics approach to explain and influence HIV prevention behaviors.

S. Linnemayr, RAND Corporation, Santa Monica, California: Using a behavioral economics approach to explain and influence HIV prevention behaviors.

M. Skovdal, University of Copenhagen, Denmark: Using theories of practice to understand HIV risk: Opportunities and challenges for new HIV prevention technologies in sub-Saharan Africa.

SESSION VI: Translating Technologies into Practice: Lessons from Other Programs

Chairperson: J. Shelton, Johns Hopkins University, Baltimore, Maryland

F. Cowan, A. Pettifor, M. Shahmanesh, D. Pillay, H. Ward

N. Mugo, J. Cleland

V. Chandra-Mouli, World Health Organization, Geneva, Switzerland: Lessons from scaling up reproductive health programs.

J. Cleland, London School of Hygiene & Tropical Medicine, London, United Kingdom: Abstinence, contraception, and condom use among single African women: What can we learn from long-term trends.

R. Ingham, University of Southampton, United Kingdom: Young people, risk, and vulnerability: The success of the English teenage pregnancy strategy and its applicability to other countries.

SESSION VII: Translating Technologies into Practice: Lessons from Three Decades of HIV Prevention

Chairperson: K. Dehne, UNAIDS, Geneva, Switzerland

N. Mugo, Kenya Medical Research Institute, Nairobi, Kenya: Holistic approach to HIV combination prevention with focus on population and geography.

M. Shahmanesh, UCL/Africa Health Research Institute, London, United Kingdom: Delivering HIV prevention: Lessons from history and key populations.

SESSION VIII: Lessons from a Country Case Study: Zimbabwe

Chairperson: G. Garnett, Bill & Melinda Gates Foundation, Seattle, Washington

F. Cowan, Liverpool School of Tropical Medicine, Harare, Zimbabwe: Lessons of HIV prevention program delivery in Zimbabwe.

K. Dehne, UNAIDS, Geneva, Switzerland: Lessons of HIV prevention program delivery in Zimbabwe.

SESSION IX: Design Workshop

Chairpersons: H. Ward, Imperial College London, United Kingdom, and **D. Pillay,** Africa Health Research Institute, Durban, South Africa

This interactive session used specific personae of high-risk individuals to map out the steps necessary for that particular person to benefit from an HIV preventive technology. At the end of the session, the various summaries were reviewed to identify key touch points needing further work.

SESSION X: Reporting and Wrap-Up

Chairpersons: H. Ward, Imperial College London, United Kingdom, and **D. Pillay,** Africa Health Research Institute, Durban, South Africa

Integrated Translational Science Center Workshop

June 12–14

FUNDED BY **National Institutes of Health/National Cancer Institute (grant awarded to D. Tuveson et al.)**

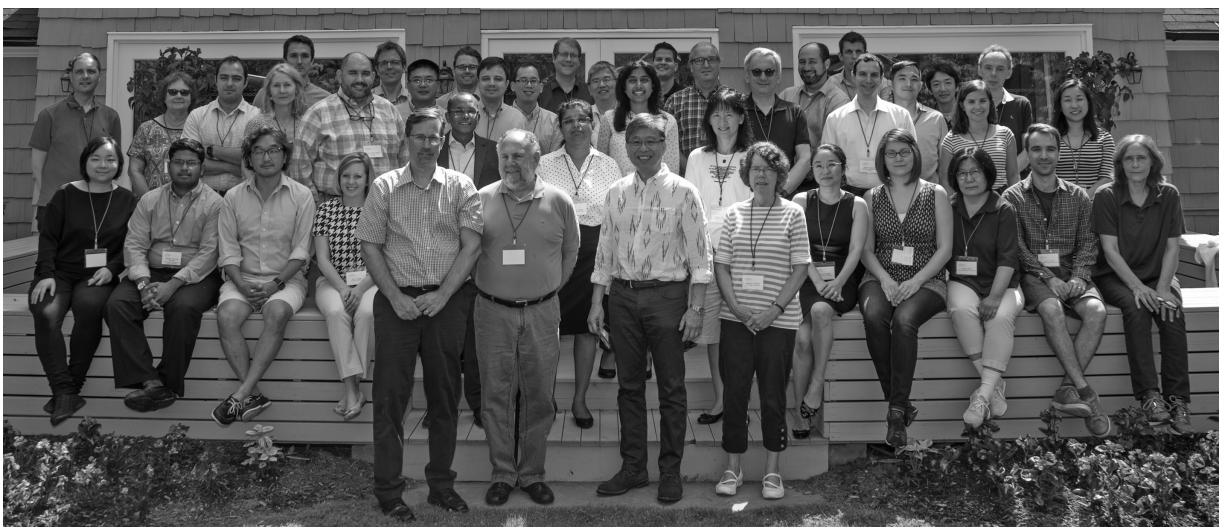
ARRANGED BY **L. Baker,** University of Michigan, Ann Arbor
L. Ellis, University of Texas, Houston
E. Liu, The Jackson Laboratory, Bar Harbor, Maine
A. Schott, University of Michigan, Ann Arbor
D. Tuveson, Cold Spring Harbor Laboratory

The Integrated Translational Science Center (ITSC) formed between SWOG, Cold Spring Harbor Laboratory (CSHL), and The Jackson Laboratory (JAX) to bridge the gap between the laboratory and the clinic by elucidating the key clinical problems and challenges in oncology that can be addressed in the laboratory, discovering new diagnostic and therapeutic approaches that can be integrated into clinical trials, and by providing a conduit for clinical trial results to be re-interpreted in the laboratory. In June, the ITSC members convened at Banbury to discuss pressing questions in cancer medicine that are best addressed in partnership with technological and basic/translational science experts at JAX and CSHL. The highly interactive meeting included talks, posters, laboratory demonstrations, and brainstorming sessions to generate ideas for collaborative projects that take advantage of both the clinical expertise of the SWOG participants and innovative science and technology provided by CSHL and JAX scientists.

Welcoming Remarks: R. Leshan, Banbury Center, Cold Spring Harbor Laboratory

Introduction to CSHL/Overview of Meeting Objectives: D. Tuveson, Cold Spring Harbor Laboratory

Introduction to Jackson Laboratory: E. Liu, The Jackson Laboratory, Bar Harbor, Maine



J. Chuang, Z. Mitri

L. Baker, D. Tuveson

Keynote Address

J. Doroshow, National Cancer Institute, Bethesda, Maryland: Evolution of NCI's Clinical Trials Networks in the Era of Precision Oncology.

Overview

E. Liu, The Jackson Lab, Bar Harbor, Maine: SWOG/CSHL/JAX ITSC Discovery Engine.

SCIENTIFIC POSTERS

Technology Presentations I

H. Reddi, The Jackson Laboratory, Bar Harbor, Maine: Clinical genomics at Jax: Technologies and capabilities.

J. Lee, Cold Spring Harbor Laboratory: Approaches to mapping gene expression signatures in space.

SWOG Investigator Presentations I

Moderator: L. Baker, University of Michigan, Ann Arbor

Panel: H. Babiker, P. Chalasani, E. Cobain, A. Danilov, A. Kirschner, J. Leonard, B. Lim, J. Markowitz

Technology Presentations II

P. Robson, The Jackson Laboratory, Bar Harbor, Maine: Cellular phenotyping tumors with single-cell technologies.

H. Tiriac, Cold Spring Harbor Laboratory: Modeling patient response in human pancreas cancer organoids.

SWOG Investigator Presentations II

Moderator: L. Baker, University of Michigan, Ann Arbor

Panel: A. Morikawa, Z. Ibrahim Mitri, A. Scott, C. Speers, A. VanderWalde, J. Carlos Varela, D. Wahl

Developing Projects

Speed Science
Pilot Presentations

Preparing Project Proposals

Moderator: L. Baker, University of Michigan, Ann Arbor

Panel: P. Robson, H. Reddi, H. Tiriac

KiiLN Foundation 2017: Bio-Entrepreneurship in NYC

July 7

FUNDED BY **Keystone for Incubating Innovation in Life Sciences Network (KiiLN), J.P. Morgan, Pfizer, Torreya Advisors, Nixon Peabody, Alston & Bird, LI Bioscience Hub**

ARRANGED BY **D. Brand,** ECHO NYC & Quicksilver Biosystems, New York, New York
N. McKnight, KiiLN & BioLabs, New York, New York

The second annual Foundation conference brought entrepreneurs, investors, industry leaders, and scientists to the Banbury Center for one day of stimulating panel sessions and networking.

Introductory Remarks: D. Brand, ECHO NYC & Quicksilver Biosystems, and **N. McKnight**, KiiLN & BioLabs New York

PANEL 1: Mining Academia

Moderator: K. Neote, Johnson & Johnson Innovation Center, Cambridge, Massachusetts;

J. Mandelbaum, Accelerator Corp, New York, New York
C. Pitt, Versant Ventures, New York, New York
T. Thornton, Northwell Ventures, New York, New York

PANEL 2: Genomics

Moderator: D. Brand, ECHO NYC & Quicksilver Biosystems, New York, New York

J. Crawford, Northwell Health, Lake Success, New York
J. Leslie, Celmatix, New York, New York
J. Pickrell, Gencove, New York, New York

PANEL 3: New Ventures

Moderator: C. Green, Pfizer, New York, New York

S. Bettigole, Quentis Therapeutics, New York, New York
J. Vacca, Highline Therapeutics, New York, New York

PANEL 4: "Dealmakers"

Moderator: A. Tinkelenberg, Torreya Advisors, New York, New York

S. Jacobson, Remedy Pharmaceuticals, New York, New York
J. Magram, Northern Biologics, New York, New York

Fireside Chat

J. Anderson, CTI LifeSciences, New York, New York

E. Schadt, Sema4, New York, New York

C. Pitt, K. Wasik, J. Pickrell, E. Berglund

S. Kongsamut, D. DeGroote, C. Green, A. Xavier

Opportunities for Reduction of Aflatoxin Contamination of Food

July 9–12

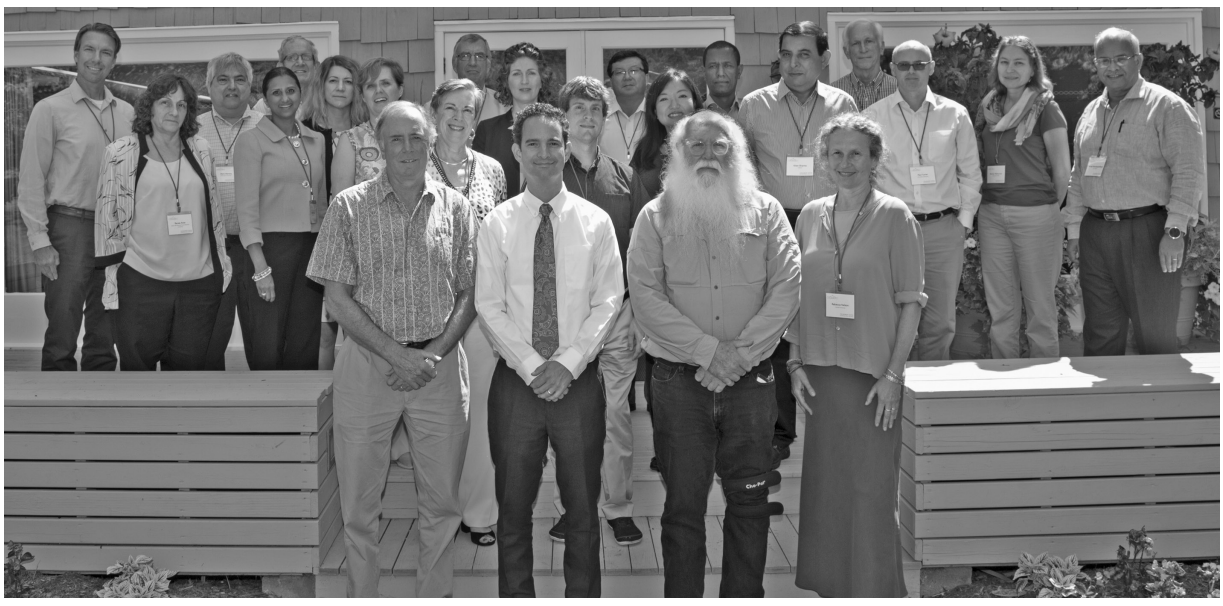
FUNDED BY **Mars, Inc., the U.S. National Science Foundation (support to R.M.), and the Cold Spring Harbor Laboratory Corporate Sponsor Program, with additional support from Kansas State University**

ARRANGED BY **J. Harvey,** Kansas State University, Manhattan
R. Michelmore, University of California, Davis
R. Nelson, Cornell University, New York

Contamination of food by mycotoxins is a worldwide problem: Estimates indicate that at least 25% of the global food supply is contaminated [UN FAO]. Aflatoxins, a type of mycotoxin produced by *Aspergillus* species, can cause liver cancer as well as a variety of ailments related to immunosuppression when adults are subjected to chronic exposure; exposure of children results in stunting. Acute exposure can result in death. Current approaches can help reduce, but not eliminate, contamination. However, recent advances in multiple areas offer new potential. This Banbury meeting reviewed current efforts to date, their limitations, and challenges. Participants assessed the new short- and longer-term opportunities enabled by advances in knowledge and technology: varieties resistant to abiotic stresses, *Aspergillus* species or insects; biocontrol; and transgenic strategies aimed at reducing pathogen growth and aflatoxin production such as host-induced gene silencing. Finally, the group considered the pathways for deployment and adoption of new intervention strategies as well as integration with efforts to reduce contamination with other mycotoxins.

Welcoming Remarks: R. Leshan, Banbury Center, Cold Spring Harbor Laboratory

Overview of Meeting Objectives: R. Michelmore, University of California, Davis



A. Ayalew, R. Michelmore

C. Woteki, A. Records

SESSION I: Current Situation, Limitations, and Challenges

Chairperson: H-Y. Shapiro, Mars, Incorporated, McLean, Virginia

H-Y. Shapiro, Mars, Incorporated, McLean, Virginia: Introduction to session themes (efforts, areas most in need, impacts on livestock, and human health).

A. Ayalew, PACA, African Union Commission, Ethiopia: Are country-led approaches the answer for coordinating efforts to win the fight against aflatoxins?

A. Bianchini, University of Nebraska, Lincoln: Mycotoxin assessment on corn production chain by small holders in Guatemala.

M. Manary, Washington University, St. Louis, Missouri: Relationship of serum aflatoxin measurements in pregnant women and newborn length.

F. Wu, Michigan State University, East Lansing: Recent findings in aflatoxin and human health: A liver cancer success story and doubts cast on stunting role.

K. Damann, Jr., Louisiana State University Agricultural Center, Baton Rouge: *Aspergillus flavus* biology and biological control of aflatoxin contamination in corn.

SESSION II: Pre- and Post-Harvest Interventions Possible in the Short Term

Chairperson: R. Michelmore, University of California, Davis

R. Bandyopadhyay, IITA, Nigeria: Biological control and other practices for aflatoxin management in Africa.

P. Cotty, Agricultural Research Service, USDA, School of Plant Sciences, University of Arizona, Tucson: Biological control: One tool reduces aflatoxins throughout the environment.

P. Ojiambo, North Carolina State University, Raleigh: Female fertility and its role in selecting effective biocontrol strains of *Aspergillus flavus.*

J. Harvey, Kansas University/Post-Harvest Loss Innovation Lab, Manhattan: Toward an integrated approach to reducing aflatoxin contamination and exposure.

R. Nelson, Cornell University, Ithaca, New York: Sorting maize at local hammer mills as part of a strategy for reducing mycotoxins in the African food system.

B. Bextine, DARPA/BTO, Arlington, Virginia: DARPA's Biological Technologies Office.

SESSION III: Conventional Breeding and Transgenic Longer-Term Interventions

Chairperson: J. Harvey, Kansas University/Post-Harvest Loss Innovation Lab, Manhattan

M. Warburton, USDA, ARS, CHPRRU, Mississippi: Translational genomics of aflatoxin accumulation resistance in maize: Can I take my lab results to the field?

W. Xu, Texas A&M University, Lubbock: Progress and challenges in breeding aflatoxin-resistant corn.

R. Michelmore, University of California, Davis: Host-induced gene silencing for disease control.

M. Schmidt, University of Arizona, BIO5 Institute, Tucson: The use of host-induced gene silencing (HIGS) to suppress aflatoxin production.

H-Y. Shapiro, J. Harvey

R. Arias, USDA-ARS-National Peanut Research Laboratory, Dawson, Georgia: RNAi-mediated control of aflatoxins in peanut.

K. Kumar Sharma, ICRISAT, India: Genetic engineering for the control of *Aspergillus flavus* infection and aflatoxin production in peanut.

SESSION IV: Translational Realities for Implementation

Chairperson: R. Nelson, Cornell University, Ithaca, New York

M. Stasiewicz, University of Illinois, Urbana-Champaign: Multispectral sorting to reduce mycotoxin levels in maize.

P. Turner, University of Maryland, College Park: The strength of exposure biomarkers in evaluating intervention strategies for mycotoxins.

C. Woteki, Iowa State University, Ames: Reducing policy barriers to lowering aflatoxin exposures.

T. Herrman, Texas A&M University, College Station: Building a public private partnership to manage aflatoxin risk through a connected and transparent marketplace that delivers aflatoxin-safe food and feed to all Africa.

N. Kazi, Humanitas Global Development, Washington D.C.: Building an enabling environment for wide-scale adoption of proven interventions that reduce mycotoxin contamination.

A. Records, USAID, Washington, D.C.: Mycotoxin mitigation: A global food security priority.

SESSION V: Next Steps

Chairpersons: J. Harvey, Kansas State University, Manhattan, **R. Nelson,** Cornell University, Ithaca, New York, **R. Michelmore,** University of California, Davis.

Neuropharmacology and Human Stem Cell Models

September 10–13

FUNDED BY **The Cold Spring Harbor Corporate Sponsor Program and Lieber Institute for Brain Development**

ARRANGED BY **N. Brandon,** AstraZeneca, Waltham, Massachusetts
Z. Cader, Oxford University, United Kingdom
S. Haggarty, Harvard Medical School and Massachusetts General Hospital, Boston

Advances in a combination of disciplines—human stem cell biology, chemical biology, and human genetics—are catalyzing new opportunities both to impact our fundamental understanding of human disease biology and to discover next-generation pharmacological agents aimed at targeting the root cause of disease. Perhaps nowhere are these advances more significant and critically needed than in the area of neurological and psychiatric diseases. Participants in this Banbury meeting, spanning neurological and psychiatric neuroscience as well as drug discovery, critically reviewed the state of human stem cell modeling as applied to the advancement of neuropharmacology. Specifically, discussion focused on the state of patient-specific bio-banking internationally; approaches to the integration of iPSC models with clinical cohorts for precision medicine, genome sequencing and deep patient phenotyping at the level of neuroimaging and neuropathology; disease-relevant stem cell-based assay development to support large-scale pharmacological and CRISPR/Cas9-based functional genomic screens; and application of stem cell models to systems pharmacology, including use of organoids and organ-on-a-chip technologies.

Welcoming Remarks: R. Leshan, Banbury Center, Cold Spring Harbor Laboratory

Introduction and Meeting Objectives: S. Haggarty, Harvard Medical School and Massachusetts General Hospital, Boston, **N. Brandon,** AstraZeneca, Waltham, Massachusetts, and **Z. Cader,** Oxford University, United Kingdom

SESSION I: Introduction and Overview

Chairperson: D. Panchision, National Institute of Mental Health, Bethesda, Maryland

D. Panchision, National Institute of Mental Health, Bethesda, Maryland: How NIMH is adapting iPSC research to functional genomics, systems neuroscience, and drug discovery.

Z. Cader, Oxford University, United Kingdom: Innovative medicines initiative platforms for iPSC drug discovery research.

A. Kaykas, Novartis Institute for Biomedical Research, Cambridge, Massachusetts: Using stem cell for neuroscience target discovery.

SESSION II: Patient Cohorts and Deep Clinical Phenotyping

Chairperson: R. Perlis, Broad Institute, Boston, Massachusetts

R. Perlis, Broad Institute, Boston, Massachusetts: Applying a large neuropsychiatric biobank to characterize Roy Perlis treatment response.

L. Studer, Memorial Sloan Kettering Cancer Center, New York: Rapid glial fates and the use of a pooled hPSC approach to identify disease phenotypes.

L. Grinberg, University of California, San Francisco: Neuropathological methods and studies on stem cell can complement each other to advance the knowledge on neurodegenerative disease.

SESSION III: Stem Cell Neurotechnology and Advanced Modeling

Chairperson: K. Fabre, AstraZeneca, Waltham, Massachusetts

K. Fabre, AstraZeneca, Waltham, Massachusetts: Microphysiological systems and inducible stem cells for drug development.

O. Brüstle, Institute of Reconstructive Neurobiology, Bonn, Germany: Programming NSCs for disease modeling and drug discovery.

E. Shusta, University of Wisconsin, Madison: Stem cell modeling of the neurovascular unit.

I. Cornella-Taracido, Merck & Company, Boston, Massachusetts: Use of quantitative, high-resolution mass spectrometry–based proteomics to enable phenotype–genotype–proteotype correlation analyses toward target and biomarker discovery in neuroscience.

SESSION IV: Alzheimer's and Related Dementia

Chairperson: F. Livesey, University of Cambridge, United Kingdom

F. Livesey, University of Cambridge, United Kingdom: Small-molecule and genetic screens to identify druggable targets in human stem cell models of dementia.

H. Inoue, Center for iPS Cell Research and Application, Kyoto, Japan: Human pluripotent stem cells in neurological drug discovery.

T. Young-Pearse, BWH and Harvard Medical School, Boston, Massachusetts: Probing heterogeneity of Alzheimer's disease using iPSCs.

SESSION V: Neurodevelopmental, Psychotic, and Mood Disorders

Chairperson: S. Haggarty, Harvard Medical School and Massachusetts General Hospital, Boston

L. Ellerby, N. Heintz

I. Cornella-Taracido, R. Livesey, H. Heimer, R. Perlis

S. Haggarty, Harvard Medical School and Massachusetts General Hospital, Boston: Advancing neuropharmacology for rare neurogenetic disorders with patient-derived stem cell models.
B. Maher, Lieber Institute for Brain Development, Baltimore, Maryland: Modeling syndromic autism spectrum disorders with patient-derived induced pluripotent stem cells.
R. Karmacharya, Harvard Medical School and Massachusetts General Hospital, Boston: Ex vivo signature of psychosis and treatment response in patient-derived neurons.

SESSION VI: Movement, Pain, and Other Disorders

Chairperson: Z. Cader, University of Oxford, United Kingdom

Z. Cader, University of Oxford, United Kingdom: Neuropharmacology and drug discovery in pain disorders.
D. Butler, Neural Stem Cell Institute, New York: Disease in a dish modeling of neurodegenerative diseases using induced pluripotent stem cells.
L. Ellerby, Buck Institute for Research on Aging, San Francisco, California: Huntington's disease: Using isogenic human HD models for target identification and drug screening.
B. Ryan, University of Oxford, United Kingdom: Identifying and exploiting phenotypes for drug discovery in Parkinson's using iPSC models.
S. Seo, Lieber Institute for Brain, Baltimore, Maryland: Spatiotemporal landscape of heterogeneity in hPSCs.

SESSION VII: Opportunities and Challenges for Drug Discovery

Chairperson: N. Brandon, AstraZeneca, Waltham, Massachusetts

D. Butler, D. Hiler

J. Erwin, Lieber Institute for Brain Development, Baltimore, Maryland: Single-cell-omic approaches to understanding cellular and genomic heterogeneity of the brain.
N. Heintz, The Rockefeller University, New York: Cell-type-specific profiling from postmortem human brain.
S. Finkbeiner, Gladstone Institutes, University of California, San Francisco: Target evaluation and small-molecule development with human cell models of neurodegenerative diseases.

SESSION VIII: Wrap-Up and Next Steps

Chairperson: N. Brandon, AstraZeneca, Waltham, Massachusetts

NLRs Sans Frontières

September 17–20

FUNDED BY **The Cold Spring Harbor Laboratory Corporate Sponsor Program and the Gordon and Betty Moore Foundation, with additional support provided by 2Blades Foundation, Burroughs Wellcome Fund, DuPont Pioneer, and Genentech.**

ARRANGED BY **J. Dangl,** HHMI/University of North Carolina, Chapel Hill
J. Jones, The Sainsbury Laboratory, Norwich, United Kingdom
R. Vance, HHMI/University of California, Berkeley

Plants and animals use intracellular proteins of the nucleotide-binding domain, leucine-rich repeat (NLR) superfamily to detect many types of microbial and viral pathogens. The specific combination of domains that define the NLR architecture likely evolved independently in each Kingdom, and the molecular mechanisms of pathogen detection by plant and animal NLRs have long been considered to be distinct. However, microbial recognition mechanisms overlap, and it is now possible to discern important trans-kingdom principles of NLR-dependent immune function. Participants at this Banbury meeting examined how a common function is achieved by NLR proteins in such diverse Kingdoms to identify features that could be useful in building new pathways through synthetic biology, whether for broadening disease defenses or for constructing new signal-response circuits. New and unpublished data shared at this meeting are expected to disrupt the community's understanding of immunity mechanisms.

SESSION I: Sensing

Chairperson: R. Vance, HHMI/University of California, Berkeley

R. Vance, HHMI/University of California, Berkeley: NLRs in plants and animals: Are there any general principles?

P. Schulze-Lefert, Max-Planck Institute for Plant Breeding Research, Cologne, Germany: Functional diversification and effector recognition mediated by a multi-allelic NLR-type disease resistance gene.

S. Shin, University of Pennsylvania, Philadelphia: Recognition of bacterial ligands by the human NAIP/NLRC4 inflammasome.

J. Jones, R. Vance

J. Dangl, B. Staskawicz

A. Daskalov, University of California, Berkeley: NLR-like proteins in fungi.

J-M. Zhou, Chinese Academy of Science, Beijing, China: Molecular links between an NLR and a PRR in *Arabidopsis*.

G. Nuñez, University of Michigan, Ann Arbor: Role of NLRs in intestinal inflammation.

E. Lien, University of Massachusetts Medical School, Worcester: Regulation of inflammasome activation by bacterial secretion systems.

SESSION II: Sensing and Signaling 1

Chairperson: J. Parker, Max-Planck Institute for Plant Breeding Research, Cologne, Germany

J. Parker, Max-Planck Institute for Plant Breeding Research, Cologne, Germany: Plant TNL nuclear receptors and defense network reprogramming.

M. Keestra-Gounder, University of Colorado, Aurora: Activation of the NOD1 and NOD2 signaling pathway.

R. Innes, Indiana University, Bloomington: Structure and function of the RPS5 NLR protein from *Arabidopsis.*

V. Hornung, Ludwig-Maximilians-University, Munich, Germany: NLRP3 inflammasome signaling in the human system.

J. Dangl, HHMI/University of North Carolina, Chapel Hill: Sensor and helper NLR function in *Arabidopsis.*

B. Staskawicz, University of California, Berkeley: Specific recognition and activation of plant NLR immune receptors.

X. Li, University of British Columbia, Vancouver, Canada: Distinct E3 ligases regulate the turnover of individual components of paired typical plant NLR immune receptors.

SESSION III: Sensing and Signaling 2

Chairperson: D. Philpott, University of Toronto, Canada

D. Philpott, University of Toronto, Canada: NOD proteins in intestinal inflammation.

J. Jones, The Sainsbury Laboratory, Norwich, United Kingdom: How the *Arabidopsis* RPS4/RRS1 immune receptor complex detects effectors and activates defense.

J. Chai, University of Cologne, Germany: Structural study on ligand recognition by an NLR protein.

F. Sutterwala, Cedars-Sinai Medical Center, Los Angeles, California: Mitochondrial regulation of NLRP3 inflammasome activation.

SESSION IV: Signaling

Chairperson: P. Dodds, CSIRO Agriculture & Food, Canberra, Australia

P. Dodds, CSIRO Agriculture & Food, Canberra, Australia: Mechanisms of plant immune receptor function in resistance to rust fungi.

A. Goverse, Wageningen University & Research, Netherlands: Distinct roles of surface regions of the CC domain in the modulation of effector-triggered immune responses by the potato resistance protein Rx1.

E. Miao, University of North Carolina, Chapel Hill: NLR-driven pyroptosis defends against intracellular pathogens.

B. Kobe, University of Queensland, Brisbane, Australia: Signaling by cooperative assembly formation by mammalian TIR domains and implications for plant NLRs.

S. Kamoun, The Sainsbury Laboratory, Norwich, United Kingdom: Evolutionary dynamics of plant NLRs: From pairs to networks.

SESSION V: Solutions

Chairperson: J. Jones, The Sainsbury Laboratory, Norwich, United Kingdom

R. Terauchi, Kyoto University, Japan: Molecular interaction and coevolution of rice-paired NLRs and *Magnaporthe oryzae*

AVRs: Similarities and differences in Pik NLR/AvR-Pik and Pii NLR/AVR-Pii interactions.

T. Kroj, INRA Montpellier, France: Unconventional integrated domains in plant NLRs provide novel insight into effector recognition and give new perspectives for the engineering of crop immune receptors.

M. Banfield, John Innes Center, Norwich, United Kingdom: Engineering an integrated domain in a plant NLR immune receptor to extend pathogen effector recognition.

P. van Esse, 2Blades Foundation, Norwich, United Kingdom: Mining NLRs from crop relatives to establish a diverse pool of disease resistance traits.

N. Krishnamurthy, DuPont Pioneer, Johnston, Iowa: Genome editing of R-genes: Opportunities and challenges.

SESSION VI: Meeting Conclusions

Metformin: Translating Biology into the Clinic

September 24–27

FUNDED BY **The Oliver Grace Chair Fund and IC-Medtech**

ARRANGED BY **N. Chandel,** Northwestern University, Chicago, Illinois
V. Stambolic, University of Toronto, Ontario Cancer Institute, Ontario, Canada

Metformin is widely used to treat patients with type 2 diabetes who exhibit high levels of circulating insulin. Recent retrospective studies have uncovered an association between metformin use and diminished tumor progression in patients suffering from different types of cancers. Given the safety of metformin along with its anticancer, anti-inflammatory, and antidiabetic effects, investigators are considering the use of metformin as the first anti-aging drug in clinical trials. This Banbury meeting convened experts to discuss the mechanisms by which metformin exerts its effects, with emphasis on translating this knowledge into the clinic.

Welcoming Remarks: R. Leshan, Banbury Center, CSHL

Introduction & Meeting Objectives: N. Chandel, Northwestern University, Chicago, Illinois
J. Watson, CSHL, Cold Spring Harbor, New York

SESSION I: AMPK

Chairperson: V. Stambolic, University of Toronto, Ontario, Canada

R. Shaw, Salk Institute for Biological Studies, La Jolla, California: Molecular dissection of Metformin action: AMPK and beyond

G. Hardie, University of Dundee, Dundee, United Kingdom: AMPK as a target for biguanides in diabetes and in cancer

D. Carling, MRC London Institute of Medical Sciences, London, United Kingdom: AMPK in health and disease: Insights using a gain-of-function mouse model

SESSION II: Mitochondria

Chairperson: S. Burgess, UT Southwestern, Dallas, Texas

N. Chandel, Northwestern University, Chicago, Illinois: Metformin target mitochondrial complex I in cancer

K. Birsoy, Rockefeller University, New York, New York: Systematic approaches to understand mitochondrial dysfunction

SESSION III: Metabolic Syndrome

Chairperson: R. Shaw, Salk Institute for Biological Studies, La Jolla, California

M. Schwab, Dr. Margarete Fischer-Bosch-Institute of Clinical Pharmacology, Stuttgart, Germany: Metformin and drug disposition: Update and future perspectives

V. Stambolic, University of Toronto, Toronto, Canada: Obesity and cancer, the insulin connection

P. Darrell Neufer, East Carolina University, Greenville, North Carolina: Viewing metformin from a mitochondrial bioenergetics perspective

S. Burgess, UT Southwestern, Dallas, Texas: Energetics of the liver during insulin resistance and NAFLD

G. Steinberg, McMaster University, Hamilton, Canada: AMPK-independent effects of metformin on metabolism: Role of GDF15 and the microbiome

SESSION IV: Cancer

Chairperson: R. Jones, McGill University, Montreal, Quebec, Canada

K. Struhl, Harvard Medical Center, Boston, Massachusetts: Anticancer effects of metformin in mouse xenografts and inducible mouse models

P. Hwang, NHLBI-NIH, Bethesda, Maryland: Using metformin to regulate aberrant mitochondrial metabolism in Li–Fraumeni syndrome

I. Romero, University of Chicago, Chicago, Illinois: Repurposing metformin for ovarian cancer treatment, including targets in the tumor microenvironment

M. Pollak, McGill University, Montreal, Quebec, Canada: Tyrosine kinase inhibitors and biguanides

G. Draetta, MD Anderson Cancer Center, Houston, Texas: Metabolic dependencies of glioblastoma stem cells

B. Zheng, MGH, Harvard Medical School, Charleston, South Carolina: Repurposing phenformin for cancer treatment

SESSION V: Aging

Chairperson: S. Budinger, Northwestern University, Feinberg School of Medicine, Chicago, Illinois

J. van Deursen, Mayo Clinic, Rochester, Minnesota: Senotherapeutic properties of metformin

A. Soukas, Harvard University, Boston, Massachusetts: Ancient metformin response pathways

SESSION VI: Immunity

Chairperson: K. Struhl, Harvard Medical School, Boston, Massachusetts

L. Morel, University of Florida, Gainesville, Florida: Metformin treatment in lupus: Evidence from mouse models and patient T cells

S. Budinger, Northwestern University, Feinberg School of Medicine, Chicago, Illinois: Metformin reduces air pollution–induced thrombosis

R. Jones, McGill University, Montreal, Quebec, Canada: Impacts of biguanides on immune cell function

SESSION VII: New Approaches in Metformin

Chairperson: M. Pollak, McGill University, Montreal, Quebec, Canada

R. Kalluri, MD Anderson Cancer Center, Houston, Texas: Perturbing vulnerable metabolic pathways using iExosomes-metformin in pancreatic cancer

I. Romeo, J.D. Watson, S. Apple, M. Schwab

G. Hardie, R. Shaw

D. Campbell, Enlibrium, La Jolla, California: Advancement of the novel biguanide, ENL069, for the treatment of cancer

SESSION VIII: Metformin Trials

Chairperson: N. Chandel, Northwestern University, Chicago, Illinois

R. Whitmer, Kaiser Permanente, UCSF, Oakland, California: Diabetes and dementia

N. Barzilai, Albert Einstein College of Medicine, New York, New York: Targeting aging with metformin (TAME)

SESSION IX: General Discussion/Meeting Wrap-Up

Chairperson: V. Stambolic, University of Toronto. Toronto, Canada

V. Stambolic, University of Toronto, Toronto, Canada: Meeting summary/review, road map for future, opportunities for funding

T. Miller, D. Carling

Cell Biology of ALS: Emerging Themes from Human Genetics

October 22–24

FUNDED BY **The Greater New York Chapter of the ALS Association**

ARRANGED BY **L. Bruijn,** ALS Association, Washington, D.C.
A. Gitler, Stanford University, California
E. Holzbaur, University of Pennsylvania, Philadelphia

A surge in the discovery of genes associated with amyotrophic lateral sclerosis (ALS) has implicated new pathways in pathogenesis of this neurological disease, including vesicular trafficking, mitochondrial quality control, autophagy, neuroinflammation, and DNA damage. In-depth knowledge of these cellular pathways will be required to target the underlying disease processes and drive therapeutic development forward. This Banbury meeting convened leaders in ALS and related fields, along with experts in the corresponding cell biology areas, to identify gaps in knowledge and to highlight opportunities for further research. Presentations and discussion were used as a springboard to identify new strategies in the fight against ALS.

Welcoming Remarks: R. Leshan, Banbury Center, Cold Spring Harbor Laboratory

Introduction and Meeting Objectives: L. Bruijn, ALS Association, Washington, D.C.

SESSION I: C9orf72

Chairperson: A. Gitler, Stanford University, California

A. Gitler, Stanford University, California: Session overview and introductory comments.

R. Baloh, Cedars-Sinai Regenerative Medical Center, Los Angeles, California; C9orf72 in neurons and microglia.

J. Ichida, University of Southern California, Los Angeles: Chemical perturbation of vesicle trafficking as a strategy to rescue C9orf72-ALS/FTD neurodegeneration.

M. Sendtner, University of Würzburg, Germany: Function of C9orf72 on actin dynamics in motor neurons and other MND-related molecules on actin dynamics in motor neurons.

R. Baloh, E. Holzbaur, P. Gopal, J. Ichida

L. Gan, V. Mootha, D. Cleveland

J. Wang, Johns Hopkins University, Baltimore, Maryland: RNA and protein homeostasis in C9orf72-linked ALS.

F-B. Gao, University of Massachusetts, Worcester: Investigating DNA damage as a therapeutic target in C9orf72-related ALS/FTD.

SESSION II: Autophagy

Chairperson: E. Holzbaur, University of Pennsylvania, Philadelphia

E. Holzbaur, University of Pennsylvania, Philadelphia: Session overview and introductory comments.

V. Gerbino, Columbia University, New York: The role of TBK1 and motor neuron autophagy in disease progression in a mouse model of ALS.

T. Lloyd, Johns Hopkins University, Baltimore, Maryland: Nucleocytoplasmic transport and autophagy in *Drosophila* models of ALS.

A. Yamamoto, Columbia University, New York: Selective autophagy and ALS.

C. Behrends, Ludwig-Maximilians-University, Munich, Germany: Roles of autophagy in ALS.

SESSION III: Mitochondria

Chairperson: V. Mootha, Harvard Medical School, Boston, Massachusetts

V. Mootha, Harvard Medical School, Boston, Massachusetts; Session overview and introductory comments.

X. Wang, Stanford University, Palo Alto, California: Regulation of mitochondrial trafficking and quality control: Implication in ALS pathogenesis.

H. McBride, McGill University, Montreal, Canada: Emerging evidence linking mitochondrial antigen presentation and neurodegeneration.

A.P. West, Texas A&M University Health Science Center, College Station: Mitochondrial control of innate immunity and inflammation.

SESSION IV-A: Other Mechanisms

Chairperson: D. Cleveland, University of California, San Diego

D. Cleveland, University of California, San Diego: Session overview and introductory comments.

S. Alberti, Max-Planck Institute of Molecular Cell Biology, Germany: RNP granules: How they form, age, and cause disease.

P. Gopal, University of Pennsylvania, Philadelphia: Dynamic, liquid-like TDP-43 RNP granules in neurons and pathologic transitions in disease.

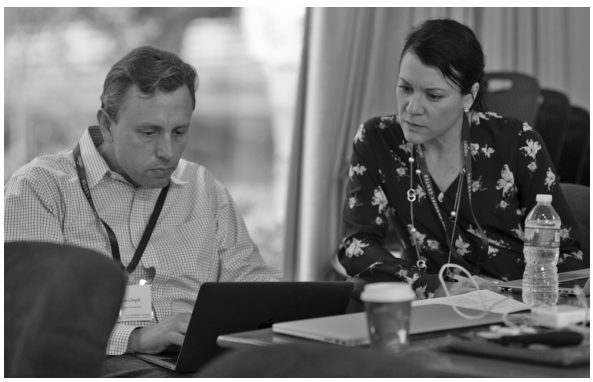
D. Milovanovic, Yale University, New Haven, Connecticut: Synaptic vesicle clusters at the nerve terminal: An example of a liquid phase?

D. Bosco, University of Massachusetts Medical Center, Worcester: Using patient-derived iPSCs to investigate mechanisms underlying FUS- and PFN1-mediated ALS.

M. Zerial, Max-Planck Institute of Molecular Cell Biology, Germany: A novel Rab5-dependent cytoprotective signaling pathway on mitochondria and its implications for ALS.

SESSION IV-B: Other cell types

Chairperson: D. Cleveland, University of California, San Diego



T. Lloyd, D. Bosco

L. Gan, Gladstone Institutes and University of California, San Francisco: Proteostasis and microglial dysfunction in neurodegenerative diseases.

K. McAvoy, Thomas Jefferson University, Philadelphia, Pennsylvania: Role of ALS astrocytes in pharmacology resistance and beyond.

J. Grutzendler, Yale University, New Haven, Connecticut: Role of glia in neurodegeneration: Evidence from optical imaging.

SESSION V: Other Diseases

Chairperson: L. Bruijin, ALS Association, Washington, D.C.

C. Sumner, Johns Hopkins University, Baltimore, Maryland: Impaired motor neuron development precedes degeneration in SMA.

K. Fischbeck, National Institute of Neurological Disorders and Stroke, Bethesda, Maryland: Therapeutics development for hereditary motor neuron disease.

SESSION VI: Meeting Summary and Roadmap for Future

Protective Immunity and Vaccines for Lyme Disease

October 29–November 1

FUNDED BY **The Steven & Alexandra Cohen Foundation**

ARRANGED BY **E. Fikrig,** Yale University, New Haven, Connecticut
S. Schutzer, Rutgers New Jersey Medical School, Newark

Lyme disease, caused by *Borrelia burgdorferi*, is the number one tick-borne disease in the United States and Eurasia. Although its cause was identified nearly 35 years ago, we still do not have a human vaccine on the market in the United States. In the past few years, our understanding of the immune response to the offending microbe has increased exponentially, as have relevant technologies. These advances now virtually assure that a safe and effective vaccine against Lyme disease can be developed and brought to market, representing a major step in decreasing cases as well as costs to the individual and society at large. Experts in Lyme disease, immunology, vaccine development, and public health convened at Banbury in October to assess the benefits and technical aspects of a Lyme vaccine.

Welcoming Remarks: R. Leshan, Director, Banbury Center, Cold Spring Harbor Laboratory

Introduction and Meeting Goals: S. Schutzer, Rutgers New Jersey Medical School, Newark

SESSION I: Overview

Chairperson: D. Rock, University of Illinois, Urbana-Champaign

R. Dattwyler, New York Medical College, Valhalla: The not-so-brief history of *Borrelia burgdorferi.*

S. Plotkin, University of Pennsylvania, Philadelphia: Historical perspective and why don't we have a Lyme vaccine?

S. Telford, Tufts University, Boston, Massachusetts: Reviving LYMErix: Rationale and strategy.

U. Lundberg, Valneva Austria GmbH, Vienna, Austria: Development of a multivalent OspA-based vaccine for prevention of Lyme borreliosis.

SESSION II: Arthropod Interactions

Chairperson: R. Dattwyler, New York Medical College

E. Fikrig, Yale University, New Haven, Connecticut: Tick immunity.

J. Benach, P. Rosa

E. Fikrig, A. Marques, J. Hovius, P. Arnaboldi

J. Pedra, University of Maryland School of Medicine, Baltimore: *Borrelia*–tick interactions.

J. Hovius, University of Amsterdam, Netherlands: Different vaccinations strategies to prevent Lyme borreliosis: Targeting *Borrelia burgdorferi* and/or the tick vector.

U. Pal, University of Maryland, College Park: Tick targets.

S. Narasimhan, Yale School of Medicine, New Haven, Connecticut: Tick immunity.

SESSION III: Determining Vaccine Efficacy in Animal Models and Humans

Chairperson: S. Schutzer, Rutgers New Jersey Medical School, Newark

S. Schutzer, Rutgers New Jersey Medical School, Newark: Detection of infection in a vaccinated individual.

P. Molloy, Imugen, Norwood, Massachusetts: The role of PCR in detection of Lyme and related TBDs.

SESSION IV: Animal Vaccines

Chairperson: W. Laegreid, University of Wyoming, Laramie

R. Marconi, Virginia Commonwealth University, Richmond: Chimeric epitope-based vaccines for tick-borne diseases.

M. Gomes-Solecki, University of Tennessee Health Science, Memphis: Oral vaccines for Lyme disease.

J. Benach, Stony Brook University, New York: Antigenic lipids of *Borrelia*.

M. Diuk-Wasser, Columbia University, New York: Part 1: Eco-epidemiological determinants for Lyme disease: Considerations for effective implementation of a Lyme vaccine.

J. Tsao, Michigan State University, East Lansing: Part 2: Eco-epidemiological determinants for Lyme disease: Considerations for effective implementation of a Lyme vaccine.

SESSION V: Bioinformatics

Chairperson: D. Rock, University of Illinois, Urbana-Champaign

S. Mitra, Icahn School of Medicine at Mount Sinai, New York: Bioinformatics applied to selection of targets and host responses.

SESSION VI: Public Health Perspectives

Chairperson: A. Marques, National Institute of Allergy and Infectious Diseases, NIH, Bethesda, Maryland

P. Mead, Centers for Disease Control and Prevention, Ft. Collins, Colorado: A public health perspective on vaccination for tick-borne diseases.

S. Plotkin

B. Backenson, New York State Department of Health, Albany: Human and tick surveillance for Lyme disease in New York: How do we know we are targeting the right culprit?

SESSION VII: Vaccine Development against Other Vectors and Targets

Chairperson: W. Laegried, University of Wyoming, Laramie

P. Arnaboldi, New York Medical College, Valhalla: TMV, a novel delivery platform for vector-borne diseases and beyond.

A. Marques, National Institute of Allergy and Infectious Diseases, NIH, Bethesda, Maryland: New challenges in Lyme disease vaccine clinical studies (STARI).

SESSION VIII: Live Attenuated versus Targeted Recombinants

Chairperson: M. Ilias, National Institute of Allergy and Infectious Diseases, NIH, Rockville, Maryland

F. Yang, Indiana University, Indianapolis: Toward development of a live attenuated Lyme disease vaccine.

P. Rosa, National Institute of Allergy and Infectious Diseases, Hamilton, Montana: Activation or neutralization of tick-borne spirochetes.

C. Cooper, USAMRIID, Ft. Detrick, Maryland: Complexity of vaccine-induced B-cell repertoires and its implications toward rationale vaccine design.

SESSION IX: Meeting Wrap-Up and Next Steps

Chairpersons: E. Fikrig, Yale University, New Haven, Connecticut, and **S. Schutzer,** Rutgers New Jersey Medical School, Newark

Lustgarten Foundation Scientific Advisory Board Meeting

November 12–14

FUNDED BY **The Lustgarten Foundation**

ARRANGED BY **D. Tuveson,** Cold Spring Harbor Laboratory
R. Vizza, Lustgarten Foundation, New York
A. Whiteley, Lustgarten Foundation, New York

Banbury was pleased to welcome back the Lustgarten Foundation for their 2017 Scientific Meeting, which provided an opportunity for the Scientific Advisory Board, as well as Foundation-supported investigators, to discuss research and strategy, evaluate performance, provide feedback for improvement, strengthen collaboration, and identify new ideas to bolster progress in the field.

Welcoming Remarks: R. Leshan, Banbury Center, Cold Spring Harbor Laboratory

Introduction and Meeting Objectives: A. Whiteley, Lustgarten Foundation, New York,
D. Tuveson, Cold Spring Harbor Laboratory, and
R. Vizza, Lustgarten Foundation, New York

PRESENTATIONS FROM TRANSLATIONAL CLINICAL PROGRAM NOMINEES

Organoid Strategy

D. Tuveson, Cold Spring Harbor Laboratory

B. Wolpin, Harvard University Medical School, Boston, Massachusetts

Project Felix: Progress and Future Aims

B. Vogelstein, Johns Hopkins University, Baltimore, Maryland

REVIEW OF RESEARCH INVESTIGATOR PROPOSALS

Distinguished Scholar Presentations

D. Fearon, Cold Spring Harbor Laboratory
R. Evans, Salk Institute for Biological Studies, La Jolla, California
B. Vogelstein, Johns Hopkins University, Baltimore, Maryland
D. Tuveson, Cold Spring Harbor Laboratory

Stand Up to Cancer Progress

D. Tuveson, Cold Spring Harbor Laboratory

Meeting Summary and Wrap-Up

A. Whiteley, Lustgarten Foundation, New York

Regulated Necrosis: Pathways and Mechanisms

November 26–29

FUNDED BY **Genentech and the Cold Spring Harbor Laboratory Corporate Sponsor Program**

ARRANGED BY **D. Green,** St. Jude Children's Research Hospital, Memphis, Tennessee
A. Linkermann, Technical University Dresden, Germany

In November, Banbury convened a second meeting on regulated necrosis, with this meeting focusing on necroptosis. After reviewing the signaling pathway, defined by RIPK3-mediated phosphorylation of MLKL and subsequent plasma membrane rupture, participants suggested mechanisms that may lead to therapeutic interference, or induction, for clinical applications. With several small-molecule inhibitors of necroptosis in clinical trials, this field has already entered the clinic. Lively discussions among the expert participants, including sharing of new and unpublished data, inspired ideas to move the field forward on biological and translational levels.

Welcoming Remarks; R. Leshan, Banbury Center, Cold Spring Harbor Laboratory

Introduction and Meeting Objectives: D. Green, St. Jude Children's Research Hospital, Memphis, Tennessee
A. Linkermann, Technical University Dresden, Germany

T. Vanden Berghe, F. Chan

H. Walczak, M. Pasparakis, J. Vince, J. Silke, P. Jost

SESSION I: TNF and Necroptosis

Chairperson: A. Oberst, University of Washington, Seattle

F. Chan, University of Massachusetts Medical School, Worcester: Climate change: Environmental control of TNF signaling.

M. Bertrand, VIB-UGent Center for Inflammation Research, Gent, Belgium: Regulation of RIPK1 life/death decisions during TNF signaling.

H. Walczak, UCL Cancer Institute, London, United Kingdom: How linear ubiquitin enables and regulates signaling: Gene activation versus different forms of cell death.

M. MacFarlane, MRC Toxicology Unit, Leicester, United Kingdom: FADD:Caspase-8 signaling complexes: Coordinated control of life/death decisions.

SESSION II: The Immune Response to Necroptosis

Chairperson: D. Vucic, Genentech, South San Francisco, California

J. Blander, Weill Cornell Medicine, New York: The many ways tissue phagocytes respond to dying cells.

M. Albert, Genentech, South San Francisco, California: Cell death and antigen cross-priming.

A. Oberst, University of Washington, Seattle: The immune consequences of necroptotic cell death in normal tissues and in cancer.

J. Vince, Walter and Eliza Hall Institute for Medical Research, Bundoora, Victoria, Australia: Live and let die: Cell death and inflammasomes.

SESSION III: Necroptosis in Viral Defense

Chairperson: E. Mocarski, Emory University School of Medicine, Atlanta, Georgia

E. Mocarski, Emory University School of Medicine, Atlanta, Georgia: Cell death, interferon, and virus-encoded suppressors of apoptosis and necroptosis.

A. Winoto, University of California, Berkeley: Necroptotic signaling pathway during Sendai virus infection.

W. Kaiser, University of Texas, San Antonio: Virus subversion of necroptosis.

J. Upton, University of Texas, Austin: Murine cytomegalovirus IE3-dependent transcription is required for DAI/ZBP1-mediated necroptosis.

SESSION IV: The Therapeutic Potential of Necroptosis

Chairperson: W. Kaiser, University of Texas, San Antonio

A. Degterev, Sackler School of Graduate Biomedical Sciences, Boston, Massachusetts: Small-molecule inhibitors of necroptosis.

A. Linkermann, Technical University Dresden, Germany: Therapeutic intervention of regulated necrosis.

T. Vanden Berghe, VIB-Ghent University, Belgium: Preclinical exploration of ferroptosis: How to induce it or block it?

P. Jost, Technical University of Munich, Germany: Necroptosis in cancer.

SESSION V: Execution Mechanisms of Necroptosis

Chairperson: T. Vanden Berghe, VIB-Ghent University, Belgium

P. Meier, Institute of Cancer Research, London, United Kingdom: Regulation of MLKL.

J. Silke, Walter and Eliza Hall Institute of Medical Research, Bundoora, Victoria, Australia: Chimeric MLKL.

A. Garcia-Saez, Universität Tübingen, Germany: Necroptosis execution by MLKL at the single-molecule level.

S. Oddo, Arizona State University, Tempe: Necroptosis activation in Alzheimer's disease.

R. Buffenstein, Calico Life Sciences LLC, South San Francisco, California: The bare essentials of healthy aging in naked mole rats.

SESSION VI: RIPK1 in Necroptosis

Chairperson: J. Vince, Walter and Eliza Hall Institute for Medical Research, Victoria, Australia

M. Pasparakis, University of Cologne, Germany: RIPK1 signaling in cell death and inflammation.

D. Vucic, Genentech, South San Francisco, California: Regulation of inflammatory cell death signaling by RIP kinases.

D. Green, St. Jude Children's Research Hospital, Memphis, Tennessee: Suppression of necroptosis by RIPK1.

SESSION VII: Meeting Summary and Wrap-Up Chairpersons: **D. Green,** St. Jude Children's Research Hospital, Memphis, Tennessee, and **A. Linkermann,** Technical University Dresden, Germany

R. Buffenstein, S. Oddo

A. Linkermann, J. Blander

Post-Traumatic Neuroinflammation: Roles in Pathogenesis of Long-Term Consequences and Repair

December 6–8

FUNDED BY **Andrea B. and Peter D. Klein**

ARRANGED BY **R. Ransohoff,** Third Rock Ventures, Boston, Massachusetts
A. Schaefer, Icahn School of Medicine at Mount Sinai, New York
D. Schafer, University of Massachusetts Medical School, Worcester

Traumatic injury of the central nervous system (CNS) is a broad term covering diverse conditions, all of which result in an incalculable toll of suffering and loss of function for the patient and a severe burden for family and caregivers. Although all traumatic conditions of the CNS occur suddenly, there is considerable evidence for an ongoing process of tissue injury that extends beyond the time of insult. Particularly relevant are the glial cells, which establish and maintain the optimal environment for neurons; glia respond to tissue injury and to neuronal dysfunction or demise, in a process termed "neuroinflammation." It is becoming apparent that the glial response to neuronal injury may be appropriate or inappropriate, helpful, harmful, or neutral—and that complex reaction programs are more the norm than the exception. This Banbury meeting brought together experts from relevant research and clinical fields to examine the effects of the glial reaction to outcomes of CNS injury and to identify promising strategies for manipulating this response to enable improved CNS function, including areas for collaboration and gaps in the overall research agenda.

Welcoming Remarks: R. Leshan, Banbury Center, Cold Spring Harbor Laboratory

Introduction and Meeting Objectives: R. Ransohoff, Third Rock Ventures, Boston, Massachusetts

E. Hughes, G. Lemke

S. Bilbo, P. Greer

SESSION I: Basic Inflammation Biology

Chairperson: A. Schaefer, Icahn School of Medicine, Mount Sinai, New York

G. Lemke, Salk Institute for Biological Studies, La Jolla, California: TAM receptors and neuroinflammation.

R. Ransohoff, Third Rock Ventures, Boston, Massachusetts: Somatic mutations in microglia and neurodegeneration.

R. Klein, Washington University School of Medicine in St. Louis, Missouri: Learning from viruses: Mechanisms of postinfectious cognitive dysfunction.

P. Greer, University of Massachusetts Medical School, Worcester: Detecting internal and external chemical cues using MS4A chemosensors.

SESSION II: Getting to Know the Glia

Chairperson: S. Bilbo, MGH Harvard Medical School, Charlestown, Massachusetts

S. Bilbo, MGH Harvard Medical School, Charlestown, Massachusetts: Brain-immune interactions in neurodevelopment: Implications for health and disease throughout the life span.

X. Piao, Harvard Medical School, Boston, Massachusetts: Glial mechanism of white matter repair.

A. Schaefer, Icahn School of Medicine at Mount Sinai, New York: Epigenetic control of regional microglia clearance activity.

E. Hughes, University of Colorado School of Medicine, Aurora: CNS injury and $NG2^+$ glial cell dynamics.

A. Mishra, Oregon Health & Science University, Portland: Regulation of cerebral blood flow in health and disease.

SESSION III: Let's Get Real: Clinical Problems

Chairperson: M. Chopp, Henry Ford Hospital, Detroit, Michigan

M. Chopp, Henry Ford Hospital, Detroit, Michigan: Stroke: Mechanisms and therapeutic approaches.

P. Gressens, Inserm, Paris, France: Integrative genomics of microglia implicates DLG4 (PSD95) in the white matter development of preterm infants.

A. McKee, Boston University School of Medicine, Massachusetts: Posttraumatic tauopathy.

S. Stukas, University of British Columbia, Vancouver, Canada: TBI in Canada: Assets and opportunities.

J. Ninkovic, Helmholtz Zentrum München & Biomedical Center of LMU Munich, Germany: Features of glial reaction to brain injury in regeneration-competent and regeneration-incompetent vertebrates.

S. Rosi, University of California, San Francisco: The role of infiltrating macrophages and complement initiation after brain injury.

SESSION IV: Injury and Repair: Models and Mechanisms

Chairperson: M. Buckwalter, Stanford University Medical Center, California

M. Buckwalter, Stanford University Medical Center, California: Adaptive immune responses and cognitive impairment after stroke.

SESSION V: Meeting Summary, Wrap-Up, and Next Steps

Chairperson: R. Ransohoff, Third Rock Ventures, Boston, Massachusetts

R. Ransohoff, Third Rock Ventures, Boston, Massachusetts: Prospects for glial-directed therapeutics to enhance outcomes of brain injury.

DNA LEARNING CENTER

DNA LEARNING CENTER EXECUTIVE DIRECTOR'S REPORT

ADMINISTRATION

Lindsay Barone
Lauren Correri
Mary Lamont
Valerie Meszaros
David Micklos
Carolyn Reid Faughnan

INSTRUCTION

Elna Carrasco
Heather Cosel-Pieper
Alison Cucco
Cristina Fernandez-Marco
Melissa Lee
Christine Marizzi
Amanda McBrien
Pauline McGlone
Erin McKechnie
Bruce Nash
Michael Okoro
Michael Paul
Sharon Pepenella
Keil Thomas
Xiaoqun Catherine Zhang

BIOMEDIA

Cornel Ghiban
Susan Lauter
Jorge Pérez de Acha
Jason Williams
Chun-hua Yang

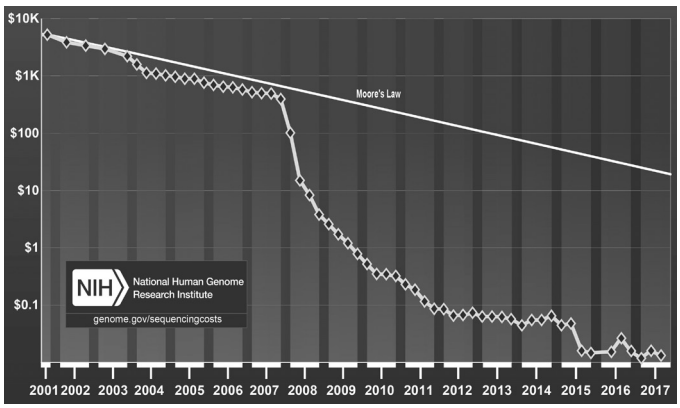
"Big data" is a common theme of much of the work at CSHL and throughout biology. This is driven, in large part, by a 50,000-fold reduction in the cost of DNA sequencing over the last decade. To put this in perspective, Moore's law states that the number of transistors on a computer chip—and, roughly, the corresponding amount of data accumulating—doubles approximately every 2 years. In comparison, the amount of raw data submitted to the National Center for Biotechnology Information (NCBI) doubled approximately every 7 months over the last decade.

Beyond DNA and RNA sequence data, big data are flooding biology from all quarters—phenotype (trait) data from agricultural field trials, patient medical records, and clinical trials; image data from microscopy, medical scanning, and museum specimens; interaction data from biochemical, cellular, physiological, and ecological systems—as well as an influx of data from translational fields such as bioengineering, materials science, and biogeography.

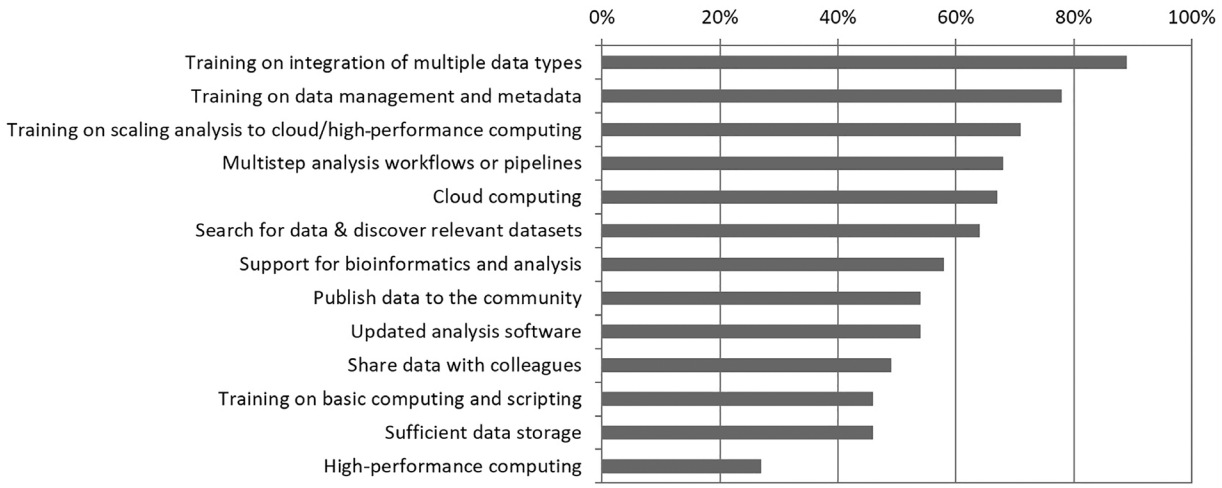
As part of our educational role in CyVerse, a major National Science Foundation (NSF) computer infrastructure for bioscience research, the DNALC has surveyed more than 1000 attendees at major professional meetings about their big data needs. Consistently, and across different conference audiences, 94% of students, faculty, and researchers said that they currently use large data sets in their research or think they will in the near future. Even so, 47% rated their bioinformatics skill level as "beginner," 35% rated themselves "intermediate," and 6% said they have never used bioinformatics tools; only 12% rate themselves "advanced." Fifty-eight percent felt their institutions do not provide all the computational resources needed for their research.

In October, we followed up on this work with a publication in *PLoS Computational Biology*, "Unmet Needs for Analyzing Biological Big Data: A Survey of 704 NSF Principal Investigators (PIs)" (*https://doi.org/10.1371/journal.pcbi.1005755*). This study gives a human face to the data revolution in biology by assessing the computational needs of active, competitive researchers who had received grants from the Biological Sciences (BIO) Directorate of the NSF.

Consistent with our previous studies, nearly 90% of BIO PIs said they are currently or will soon be analyzing large data sets. Not surprisingly, biological sequences topped the list of data types used by the PIs, followed by images, phenotype, ecological, and microscopy data.



Sequencing costs per raw megabase of DNA sequence



Unmet needs of BIO PIs—percentage who said a need was not met by their institution.

The BIO PIs considered a range of computational needs important to their work—including high-performance computing (HPC), bioinformatics support, multistep workflows, updated analysis software, and the ability to store, share, and publish data. However, a majority of PIs—across bioinformatics/other disciplines, large/small research groups, and four NSF BIO programs—said their institutions are not meeting nine of 13 needs.

Hardware is not the problem. BIO PIs ranked availability of data storage and HPC lowest on their list of unmet needs. This provides strong evidence that the NSF and individual universities have succeeded in developing a broadly available infrastructure to support data-driven biology. Training on integration of multiple data types (89%), on data management and metadata (78%), and on scaling analysis to cloud/HP computing (71%) were the three greatest unmet needs. BIO PIs saw training as the most important factor limiting their ability to use big data and to integrate data obtained from different kinds of experiments and computational platforms. This sort of integration will be required for a deeper understanding of "the rules of life"—notably genotype–environment–phenotype interactions that are essential to predicting how agricultural plants and animals can adapt to changing climates.

Thus, our study identified a growing gap between the accumulation of many kinds of data—and researchers' knowledge about how to use it effectively. Funding agencies need to recognize that significant new investments in training are now required to make best use of the biological data infrastructures they have helped establish over the last decade.

DNA Barcoding and Microbiomes

The DNALC administers three programs that demonstrate different models for using DNA barcoding in high school research. *Barcode Long Island (BLI),* funded by the National Institutes of Health (NIH), involves students in collaborative "campaigns" to compare biodiversity across Long Island. The *Urban Barcode Project (UBP),* funded by the Thompson Family Foundation, and *Urban Barcode Research Program (UBRP),* funded by matching grants from the Pinkerton Foundation and Simons Foundation, involve students in independent projects to explore biodiversity in New York City (NYC). *BLI* and *UBP* students are mentored by classroom science teachers, whereas *UBRP* students are mentored by scientists from different NYC research institutions.

We moved our DNA barcoding work into the realm of big data when we received a supplemental grant from the NIH Big Data to Knowledge Program to adapt microbiome research for high

school students. Whereas DNA barcoding examines a DNA sequence from a single organism, studying microbiomes uses next-generation sequencing (NGS) to analyze hundreds of thousands of DNA sequences representing complex mixtures of microbes found in environmental samples. In a typical experiment, DNA is isolated from water or soil samples taken from different environmental locations, polymerase chain reaction (PCR) is used to amplify a variable region of the 16S ribosomal RNA gene, and NGS reads identify the variety and abundance of microbial species from different locations.

During the summer, twelve *BLI* faculty mentors received 8 days of training in microbiome research, covering project design and sample collection, biochemistry, and data analysis. With support, these mentors guided 61 students on 22 teams as they produced and analyzed more than 11 billion nucleotides of sequence from DNA isolated from 220 samples. To facilitate analysis using the command-line program QIIME (Quantitative Insights into Microbial Ecology), we introduced faculty to Jupyter Notebooks, which combines live computer code with explanatory text and visualizations to create reusable workflows. Highlights of the preliminary analyses included finding microbes related to eel grass wasting disease in the Great South Bay, identifying human pathogens in Long Island ticks, and finding arsenic-tolerant microbes in cemeteries of the Civil War era, when arsenic was used for embalming.

Joslynn Lee, Bruce Nash, and Sharon Pepenella (*left* to *right* at *far right* of image) lead students through microbiome analyses using Jupyter notebooks.

In addition to microbiome research, the 2017 *BLI* program included 251 students doing traditional DNA barcoding projects. *BLI* teams represented 30 public and private high schools from Suffolk, Nassau, and Queens; 17% of participants were African American, Latino, and Native American. To complete their research in time for the annual research symposium, 130 students from 47 teams attended 13 open lab sessions held at the Dolan DNALC, DNALC *West*, Stony Brook University (SBU), or Brookhaven National Laboratory (BNL). Teams processed more than 1400 samples, resulting in more than 1600 sequencing reads. Students published 24 sequences in GenBank, including 11 novel barcode sequences.

Eighty-five DNA barcoding and 21 microbiome projects were presented at the annual *BLI* research symposium on June 7, 2017 at CSHL and included biodiversity studies of plants, invertebrates, fungi, algae, and lichens, plus microbiome studies of water, soil, invertebrates, plants, and vectors of disease. Dr. Jeremy Seto of the New York City College of Technology gave the keynote

The *Barcode Long Island* Symposium was held on the CSHL campus in the Nicholls Biondi Hall, Bush Lecture Hall, and Grace Auditorium in June.

(*Left*) The New York Academy of Medicine provided a beautiful space for *UBP* and *UBRP* students to present research posters. (*Right*) Jesse Ausubel (second from *left*) speaks with students during the symposium.

address on microbiome analyses. *BLI* students received a number of awards at the Long Island Science Congress competition. Nathalia Reis from William Floyd High School, under the mentorship of teacher Victoria Hernandez, was also awarded first place in the Junior Biological Science Division at the NYS Science and Technology Entry Program (STEP), and Elizabeth Scianno of William Floyd High School was invited to present her research at the NIH Citizen Science Symposium in Bethesda, Maryland.

The 2017 *UBP* and *UBRP* programs included 178 students, 27% of whom were African American, Latino, or Native American, and represented 28 public and 16 private high schools from NYC. *UBP* students made ample use of DNALC resources: 68 students from 23 teams attended open lab sessions at *Harlem DNA Lab* or Genspace, whereas 98 students from 34 teams borrowed equipment footlockers for use at school. Teams collected and processed more than 1,170 samples for DNA sequencing, resulting in more than 1,900 single sequences and one million NGS reads—and produced 10 new GenBank entries. The annual research symposium on May 25, 2017 at the New York Academy of Medicine showcased 62 projects and included a keynote speech by Jesse H. Ausubel, of Rockefeller University, on the history of DNA barcoding. The winning *UBRP* team built a system of reference DNA barcodes for fish species in the Bronx River and identified one sample as American plaice (*Hippoglossoides platessoides*), a species that is not known to occur in the Bronx River. The winners for *UBP* assessed plant biodiversity of the Thain Family Forest in the Bronx; about a quarter of barcoded roots found in soil samples were from the invasive Eurasian grass *Poa annua*.

This year, 221 students across all three barcoding programs (*BLI, UBP,* and *UBRP*) took surveys as a part of our ongoing effort to monitor the impact of participation in science research. Participants were asked about their experiences in the programs, how much they had learned, and how they felt about science. The students were overwhelmingly proud of the research they had done (88.7%) and felt that the approach to problem-solving they learned through DNA barcoding research would be helpful in future science courses (83.7%) and careers (76.8%). They also reported that research participation had altered their desire to pursue science in the future, with two-thirds indicating they were more interested in studying science in the future (67.9%) and, specifically, studying biology (67.2%). Overall, our results suggest that DNA barcoding effectively demystifies the process of science research and encourages students to continue pursuing science as a potential career path.

CyVerse and MaizeCODE Projects

The DNALC leads the outreach effort for these two major NSF research projects. CyVerse (Cyber Universe) is a 10-year, $100 million collaboration between CSHL, the University of Arizona, and

the Texas Advanced Computing Center to develop a national computer infrastructure to support biological research. The mission of CyVerse Education, Outreach, and Training (EOT) is to empower biologists at all levels of professional preparation to effectively use computer infrastructure to unlock the value of biological data for research and teaching. The annual training effort reached 268 researchers and educators at 10 workshops in the United States and the United Kingdom, whereas monthly introductory webinars reached 78 virtual attendees. As part of our commitment to diversity, CyVerse also supported workshops at the annual meeting of the *Society for the Advancement of Chicanos/Hispanics and Native Americans in Science* (*SACNAS*). *DNA Subway*, a set of classroom-friendly interfaces to bioinformatics workflows, drew 61,286 visitors—who created 36,696 new projects, viewed 1.08 million pages, and spent an average of 19 minutes using the site. There were 5905 new *DNA Subway* registrations, which accounted for more than half of new CyVerse accounts. Visiting bioinformatics researcher Jorge Pérez de Acha made significant updates to *DNA Subway's* Green Line for RNA-Seq analysis and began development of a new Purple Line for metabarcode analysis. Both are to be released in 2018.

MaizeCODE is a collaborative effort between researchers at CSHL and New York University to create a comprehensive reference encyclopedia of DNA sequences that control transcription in maize and teosinte. MaizeCODE data will provide an important resource for breeders and plant scientists to improve crop traits such as disease resistance, drought tolerance, and yield by providing high-quality genome sequences paired with diverse molecular data. To prepare for the outreach program, we conducted a survey of participants at the Maize Genetics Conference in St. Louis to clarify how the maize community intends to use the MaizeCODE data sets in research and education. Notably, 60% of participants ($n = 112$) said they would participate in annotating a gene family of interest or evaluating community annotations and ~40% said they would annotate a gene family for a class project.

Against this backdrop, we began to redevelop the *DNA Subway* Red Line for the community annotation of the maize genome. Crucially, we upgraded to WebApollo, which works transparently with JBrowse and is a great improvement over the previous desktop version. We tested this workflow at a 2-day "Annotation Jamboree" conducted in December at CSHL with graduate students representing eight maize and sorghum research groups. The participants worked in groups to check the accuracy of the automated gene models in the maize reference genome (B73v4) from five gene families: *PIN, GH3, ABC, TCP,* and *ORC*. Suspicious annotations were identified based on low scores on annotation edit distance (AED) and quality index (QI), which gauge how well a gene model is supported by available RNA evidence. About 20% of genes needed manual

(*Left*) CSHL CyVerse and MaizeCODE staff led a half-day "Bioinformatics Workshop for Plant Genomics" prior to CSHL's Plant Genomes & Biotechnology meeting in December. (*Right*) Dr. Monica Munoz-Torres, Phoenix Bioinformatics and former Project Manager of the Apollo project, gave an introduction on the importance of community curation efforts and an in-depth demonstration of Apollo's capabilities.

curation, whether they were flagged or not. Improvements included setting exon boundaries, identifying noncanonical splice sites, and adding missing exons or UTRs. Approximately 70% of genes tagged by AED/QI12 metrics required corrections to the primary gene model. Disagreements between annotators were few, and mainly involved isoforms or UTR length. Interestingly, three of the flagged genes had been correctly annotated in the earlier version of the maize genome (B73v3), showing that this method readily catches discrepancies between genome versions.

Biotechnology in American Schools: Then and Now

During the summer we received a discretionary grant from NSF program officer Celeste Carter to conduct a longitudinal survey of high school biology teachers. As part of an early grant from the NSF's Advanced Technological Education (ATE) Program, we conducted a nationwide survey of 4,100 high school biology teachers. This purposive sample took a snapshot of biotechnology/molecular genetics instruction in U.S. high schools in 1998. The new survey will repeat that study to see where biotechnology instruction stands two decades later. The survey will compare lab instruction and student exposures to four major techniques of biotechnology and molecular genetics that were measured in the original survey: bacterial transformation, DNA restriction analysis, PCR, and DNA sequencing. Responses will be compared across a range of teaching constraints, demographic, and behavioral variables in the original study.

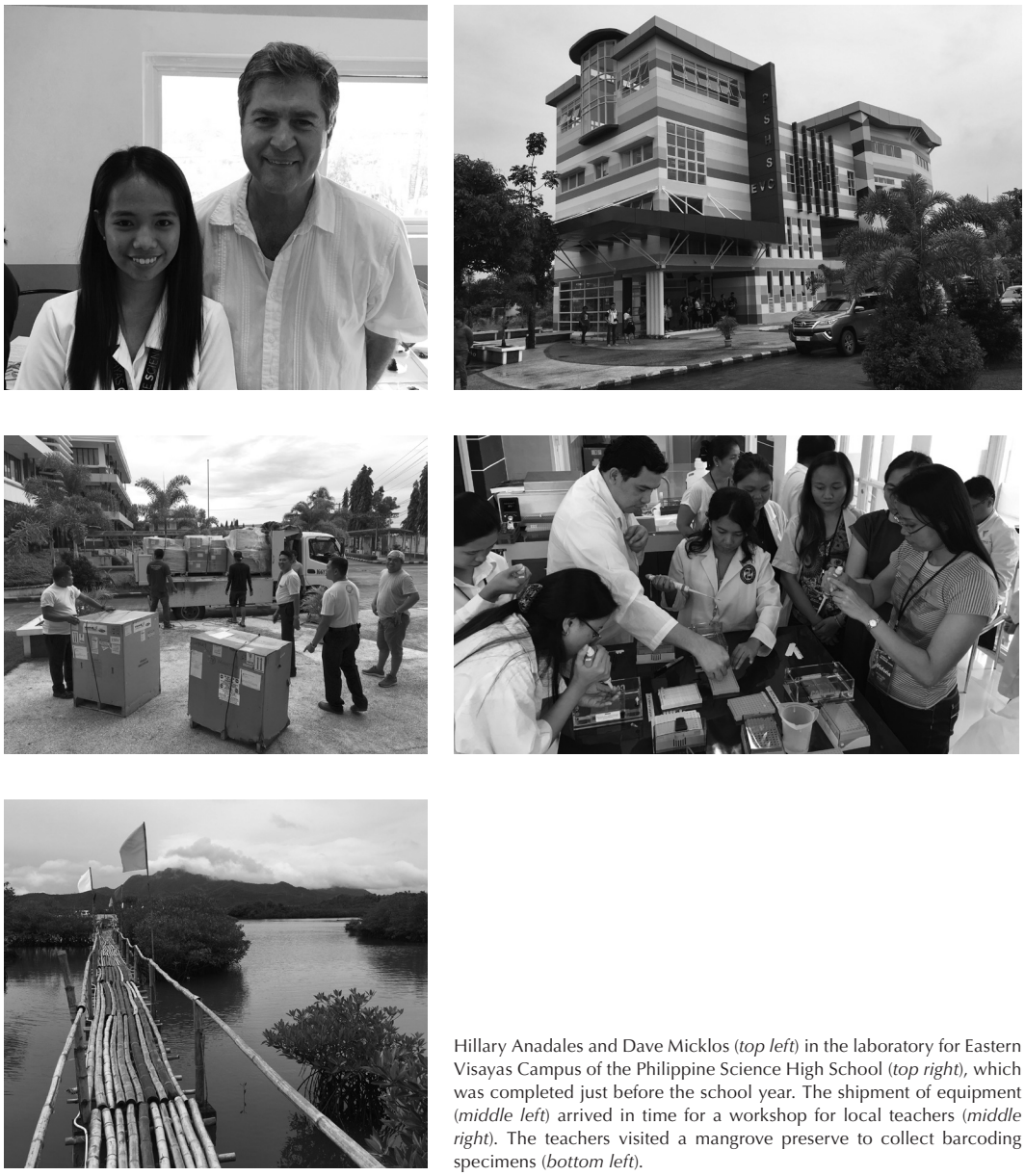
To increase the historical base, we will add data from at least 600 teachers trained at week-long workshops conducted from 1987 to 1996. These surveys had identical or comparable questions to the nationwide survey, but also included semantic differentials—which provide a unique way to measure attitudes. Teachers rated their reactions to "recombinant DNA, biotechnology" and "myself as a biology teacher" on 18 scales of polar adjectives ("important–unimportant," "messy–neat," "dangerous–safe," etc.). Responses cluster in dimensions of potency, activity, and evaluation—providing a validated way to track shifts in teacher attitudes. Participants at each time point will be sorted into matched groups based on school and teacher demographics.

Our 1998 study showed that the early adoption of biotechnology lab instruction was concentrated in schools located in high-wealth suburban zip codes. It is important to see if, today, hands-on biotech instruction has broadened to include poorer and more diverse urban populations. We also hope to correlate changes in teacher attitudes and behavior with funding and educational trends over this period of time—including NSF's shift away from teacher training institutes and the increased standardized testing under *No Child Left Behind*. In this way, the study results can provide insights for educational policy.

The DNALC possesses unique data sets from the 1980s and 1990s. Many of the teachers represented in those surveys have since retired. So this longitudinal survey offers the rare opportunity to compare two generations of biology teachers. The previous generation pioneered the introduction of new labs that illustrated the basic concepts of molecular genetic manipulation, mostly using the microbial model *Escherichia coli*. The current generation of teachers faces the challenge of bringing biology instruction into the age of whole-genome analysis of humans and other organisms. This survey will help us see how far biology instruction has come over the last 30 years, and where it needs to point for the next 30 years.

Breakthrough Junior Challenge

In 2015, we began a collaboration with the Breakthrough Junior Challenge, a global competition in which precollege students produce short videos explaining an important science concept. Funded by Mark Zuckerberg, Priscilla Chan, and Yuri and Julia Milner, the Junior Challenge is a complement to the prestigious Breakthrough Prize, designed to inspire creative thinking about fundamental concepts in the life sciences, physics, or mathematics. Winning students



Hillary Anadales and Dave Micklos (*top left*) in the laboratory for Eastern Visayas Campus of the Philippine Science High School (*top right*), which was completed just before the school year. The shipment of equipment (*middle left*) arrived in time for a workshop for local teachers (*middle right*). The teachers visited a mangrove preserve to collect barcoding specimens (*bottom left*).

receive a $250,000 scholarship, and their school receives $100,000 to renovate and equip a state-of-the-art *Breakthrough Science Lab*. The DNALC works with the school principal and science faculty to design the lab and acts as purchasing agent to procure quality equipment and supplies at CSHL's negotiated rates. DNALC staff members then provide on-site training for school faculty to implement labs from its *DNA Science*, *Genome Science*, and *DNA Barcoding 101* curricula.

The first two *Breakthrough Science Labs* were completed in 2017: one for 2015 winner Ryan Chester at North Royalton High School, Ohio, and one for 2016 winner Hillary Andales at the Philippine Science High School of Eastern Visayas. Although in totally different settings, each lab follows the blueprint of DNALC's teaching labs—including our signature lab bench. Establishing the Philippines lab was especially significant, because the Eastern Visayas campus was severely

damaged by the first landfall of Super Typhoon Yolanda (Haiyan) in November 2013. The 13-foot storm surge obliterated large sections of the local city, Tacloban, and was responsible for 2,300 deaths. The Science High School provided temporary housing for 3,000 displaced persons, and students had to be relocated to other campuses.

Working with contractors and shipping agents to complete a lab over a distance of 13,000 miles proved challenging. Erin McKechnie worked with Michael Marchesiello and Liz Janow of the CSHL Procurement Department to ship several pallets full of equipment and supplies to the school in time for a teacher training workshop scheduled for November. However, when Dave Micklos arrived in the Philippines the weekend before the workshop, the equipment was lost in transit, and two importers were arguing about which Filipino airport had impounded the goods. There seemed little chance that the equipment would arrive in time for the workshop to begin on Monday. Miraculously, the entire shipment arrived at the Science High School on Sunday at 4:30 pm, filling an entire flatbed truck. Then commenced a mad dash to unload the equipment and ready the lab for the next day.

The workshop was a great success, with teachers representing most of the country's network of science high schools. In addition to labs on bacterial and human genetics, the teachers also used DNA barcoding to conduct a mini-biodiversity study of a mangrove nature preserve. We hope to set up a footlocker system for sharing equipment among the science high schools to mount a national DNA barcoding campaign. This effort will be strengthened by the fact that Hillary Andales also won the 2017 Breakthrough Junior Challenge, providing the Philippine Science High School with a second lab and additional teacher training to focus on DNA barcoding.

Licensed DNALC Centers

Cold Spring Harbor Asia DNALC is a collaboration between CSHL and the Suzhou Industrial Park (SIP) with the mission of bringing interactive, American-style biology instruction to Chinese students. Started as a nongovernmental organization (NGO) in 2015, the project faced no shortage of challenges—including managing lab logistics 7,000 miles away and navigating the complexities of the Chinese educational system. The project made great strides in 2017; under the leadership of Education Director Dr. Jessica Talamas and her Chinese counterpart, Director Melissa Du,

Jason Williams (*left*) leads students through an exercise in the BioCoding camp, which introduces basic skills in bioinformatics.

Melissa Du participated in the *Saturday DNA!* "Dust Away Crime: The Truth about Fingerprints" session during her visit in October.

Jessica Talamas (*right*) oversees students at the Cold Spring Harbor Asia DNA Learning Center in Suzhou.

lab instruction expanded to reach 2,961 students. To accomplish this, a core set of DNALC labs was adapted for Chinese students and translated into Mandarin, and abbreviated versions of *DNA Science* and *Genome Science* books were created and updated for use in 5-day summer camps. Extensive work was also done to adapt and troubleshoot protocols, as many American reagents are unavailable or prohibitively expensive in China.

DNALC at Beijing No. 166 School was founded in 2014 with funding from the Dongcheng Education Commission. Since that time about 800 students and 500 teachers have participated in workshops and training courses in New York or Beijing. In 2017 we signed a new 3-year contract, with the objective of establishing *Barcode Beijing* to involve middle and high school students in independent research projects studying biodiversity in Beijing. In winter 2017, 31 Beijing 166 students took *DNA Science* and *DNA Barcoding* courses in New York. In summer, 48 students came to CSHL and completed labs from the middle school camps *Fun with DNA*, *World of Enzymes*, *Green Genes,* and *Forensic Detectives.* Students also met Dr. James Watson and attended science seminars given by CSHL scientists. DNALC staff also conducted *Human Genome Science*, *Genome Science,* and *DNA Barcoding* workshops for 122 students and 70 teachers in Beijing.

In 2017, 2,660 students participated in programs sponsored by the *DNALC* at the *University of Notre Dame*, under the leadership of Dr. Amy Stark. School-year instruction included field trips to the DNALC, in-school programs, and staff participation in local, regional, and state-level science fairs. Students from around the country also attended Notre Dame summer

DNALC educator Katie McAuley visited Suzhou in July to instruct *Fun with DNA* and *World of Enzymes* camps.

The DNALC Asia teaching labs feature the same unique lab benches designed for the CSHL DNALC.

Beijing 166 School summer camp students pose with Dr. Watson (*top row, center*) following his talk.

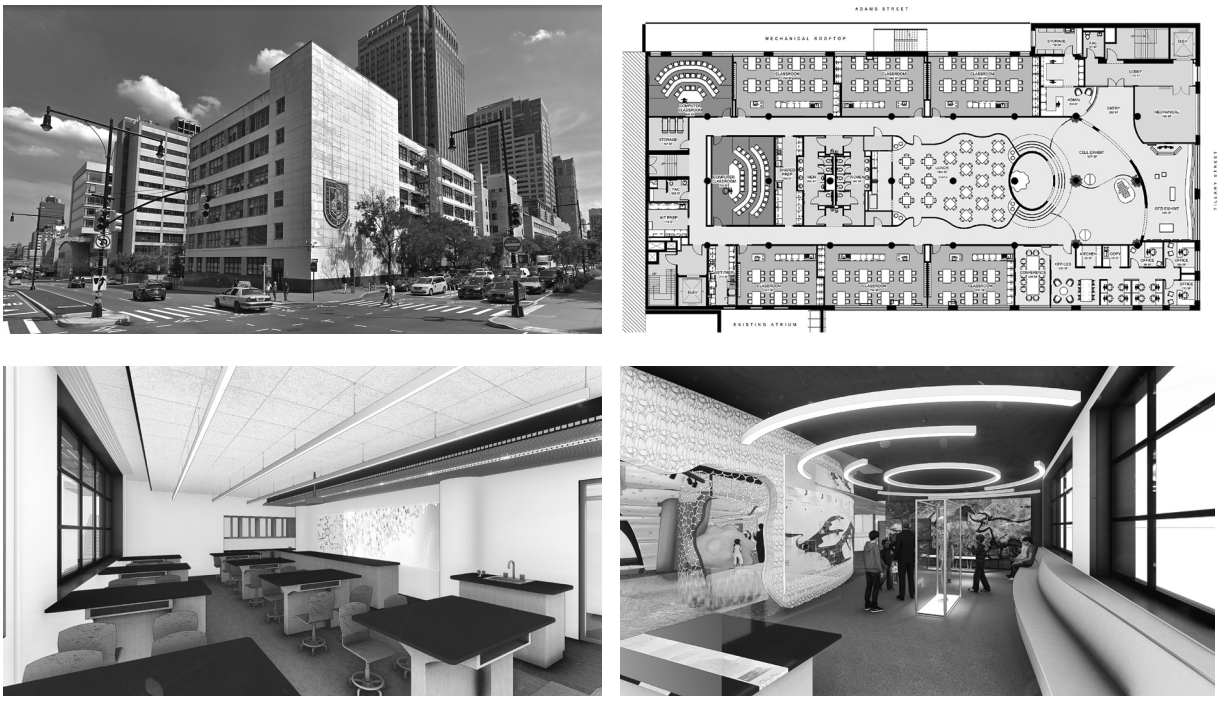
programs that included week-long residential and day camps, single-day DNA workshops, and a new 2-day introductory course for elementary students.

DNALC @ City Tech

The project to develop a permanent DNALC presence in New York City made great strides in 2017. During the summer, the Boards of Trustees of CSHL and the City University of New York (CUNY) approved a memorandum of understanding that included the terms of a 30-year, no-cost lease of the second-floor space at CUNY College of Technology (City Tech). With this approval in hand, Centerbrook Architects and Planners began developing advanced conceptual designs for the 17,500 square foot space, which occupies the entire second floor of the City Tech building on the corner of Tillary and Adams Streets in downtown Brooklyn. The plan includes:

- six windowed lab classrooms around the building perimeter,
- an internal "core" with exhibit, lunchroom, and rest rooms,
- two computer labs,
- three prep labs, and
- administrative offices.

When open in 2019, we anticipate the DNALC @ City Tech will run a full schedule of activities like those currently available at the Dolan DNALC—academic year lab field trips, summer DNA camps, and weekend family activities. The no-cost lease means that all money raised for operating costs or endowment will go toward programs. Our business plan provides scholarships for at least 50% of students taking academic year field trips. As an institution that offers both 4-year and 2-year degrees, City Tech will benefit from the DNALC's grant-funded programs for college faculty and students. This includes our active involvement in the NSF Advanced Technological Education (ATE) Program, which focuses on improving science instruction at community colleges. As part of our commitment to collegiate education, two labs will be dedicated to course-based research—allowing at least 500 CUNY students per year to make use of the DNALC's integrated programs in DNA barcoding, microbiomes, RNA sequence analysis, and genome annotation.



Centerbrook Architects and Planners worked closely with DNALC and CSHL facilities staff to develop plans for the DNALC @ City Tech: (*top left*) planned site, (*top right*) floor plan, (*bottom left*) lab, (*bottom right*) interior.

Lab Instruction and Public Outreach

In 2017, 21,007 students attended lab field trips at our three facilities: Dolan DNA Learning Center, DNALC *West*, and *Harlem DNA Lab*. In-school lab instruction was provided for 9,358 students, and 1,411 students attended week-long winter and summer camps. An additional 2,054 students used footlocker kits, including 360 conducting independent research through *UBP*, *UBRP*, or *BLI*.

Grants from Bank of America, TEVA Pharmaceuticals, and the National Grid Foundation provided field trip scholarships for 1,850 students from Long Island public schools—including Brentwood, Uniondale, William Floyd, Central Islip, Roosevelt, and Valley Stream. The William Townsend Porter Foundation provided scholarships for 1,890 students at *Harlem DNA Lab* (65% of total students). Forty students from the New World Prep Charter School received scholarships for *Fun with DNA* and *World of Enzymes* summer camps held on the school campus in Staten Island, and 15 students from IS59 in Queens participated in an ongoing collaboration with Northwell Health at DNALC *West*.

The Partner Member Program continued to provide custom science sequences and advanced electives for eight schools (primarily independent) in the tristate region. New member Grace Church School offered a 3-week summer program that combined *Genome Science* with DNA barcoding to survey biodiversity of the Hudson Estuary. At Marymount School of New York, middle school students received more inquiry-based instruction, and high school students continued to do *Genome Science* and DNA barcoding experiments as key parts of a molecular genetics elective. A barcoding

For the partial eclipse on August 21, 2017, DNALC educators provided summer campers with eclipse-viewing glasses so they could step out of the lab classrooms at peak viewing time.

In the *Being Human* summer camp, students use anatomical, biochemical, fossil, and cultural evidence to build a picture of how humans evolved.

team from Sacred Heart Academy (formerly Convent of the Sacred Heart) joined the *Billion Oyster Project*, which is devoted to restoring the oyster population in the Hudson Estuary. Eighth grade research students at Lycée Français de New York continued to use DNA barcoding to detect food fraud and survey the biodiversity of NYC. Riverdale Country School offered ninth and 11th grade labs on gel electrophoresis and PCR. Chapin School added an agar art component to an existing microbiome study. St. David's School continued fifth and eighth grade programs, including a *Forensic Detectives* summer camp. As part of our continued collaboration with Long Island schools, two classes were co-taught by DNALC staff and school faculty. Students taking molecular and genomic biology electives at Cold Spring Harbor High School and St. Dominic High School received daily lab instruction from DNALC staff members.

During the year, our Ötzi the Iceman exhibition drew 6,100 visitors. Some were students performing activities developed by our instructors; some were the general public taking self-guided tours. An additional 170 visitors registered for interpretive tours. Many were drawn to the exhibit after viewing the NOVA special *Iceman Reborn* that documents the creation of the 3D replica of the Ötzi mummy in our exhibit. We developed a new middle school lab to coordinate with the exhibit. In this ancient forensic mystery, students identify types of pollen found in different parts of Ötzi's intestinal tract to deduce where he spent the last few days of his life. This lab is being developed as a commercial kit, under our long-term contract with Carolina Biological Supply Company (CBSC). We are also upgrading kits to meet Advanced Placement Biology standards to encourage students to test hypotheses about the mechanics of key labs on bacterial transformation and restriction analysis. Kits developed under the CBSC are used by approximately 300,000 students annually, significantly extending the DNALC's reach.

Eight *Saturday DNA!* sessions drew 247 participants. Participants learned about lactose intolerance, identified an inherited transposon in their own DNA, compared their own mitochondrial DNA to modern humans and ancient relatives, and investigated the forces behind evolutionary change. They designed experiments to extract DNA from a variety of fruits and vegetables and purified a fluorescent protein from genetically engineered bacteria. Two forensics sessions included fingerprint analysis and solving the *Mystery of Anastasia*.

Several new programs were presented in collaboration with the CSHL Women in Science and Engineering (WiSE) group. "Be WiSE about Neuroscience" included a talk and panel discussion with CSHL researcher Dr. Camila dos Santos for parents, and hands-on activities with TENS (transcutaneous electrical nerve) stimulators for children. A WiSE *Fun with DNA* summer camp for girls included special activities presented by enthusiastic young role models pursuing careers in science—including a lab on herd immunity, Bollywood-style dancing to simulate mitosis, and an explanation of CRISPR gene editing technologies. Parent Day included student presentations and walking tours of CSHL.

Cold Spring Harbor High School Molecular and Genomic Biology course participants visit Ötzi.

We partnered with the American Society of Microbiology (ASM) to host several Agar Art workshops for the general public. Pieces completed at these workshops were automatically eligible for submission to the ASM Agar Art 2017 contest. Jenny Xu, a student from Chapin School in NYC, won this year's DNALC partner contest and was featured at the ASM Microbe conference for her submission "Blooming Microbes."

Our collaboration with the Watson School of Biological Sciences continued to train graduate students in skills needed to communicate science to almost any audience. Graduate students work with DNALC instructors to complete 12 half-day teaching sessions, which are designed to prepare graduate students to quickly assess an audience and customize a presentation accordingly. After completing both middle and high school rotations, graduate students chose three elective workshops to implement the skills they developed.

Students and parents dust for fingerprints during a *Saturday DNA!* session.

BioMedia Visitation and Projects

In 2017, 6.4 million visitors accessed our suite of multimedia resources. Google Analytics counted 4,661,735 visits to DNALC websites, our YouTube videos received 942,861 views, and the *3D Brain*, *Weed to Wonder*, and *Gene Screen* smartphone/tablet apps were downloaded 816,669 times. In-app purchases of 3D Brain HQ netted $7,687 for the year. Visitation to DNALC websites has decreased, partly because of changing technology and partly because of content "aging out" (several of our websites were launched more than 15 years ago). We are forming new strategies to cope with these changes.

In developing our first interactive website in 1998, *DNA from the Beginning (DNAftB)*, we were among the first groups to make extensive use of Macromedia Flash, a cutting-edge media integration program now owned by Adobe. Flash technology allowed us to create additional innovative sites with major federal and private funding, including *Inside Cancer* and *DNA Interactive*. Although Flash allowed animations and video to run smoothly on web browsers, it began to fall out of favor with the rise of smartphones and tablets—which do not support the technology. In several years, Flash will no longer be supported in browsers on computers.

So, we now face the challenge of recreating our sites using newer HTML5 technology. This entails editing each Flash animation, testing interactivity programming, and repackaging the new HTML5 animation in a reconfigured website. We used this conversion process in 2017 to successfully update *DNAftB,* which has 41 "concepts" or chapters—each with introduction, animation, gallery, videos, biographies, problem animation, and links. The animation and problem sections used Flash, so a total of 82 interactive files required conversion. After several months of testing and debugging, the smart device-friendly site launched in November.

Joe Rossano's *Conservation from Here* exhibition at the Oyster Bay Historical Society featured images of DNA barcoding specimens collected by students attending camps at the DNALC.

The *BioMedia* Group continues to support the initiatives of the DNALC through web and print design, photography, videography, exhibition development, and lab classroom layout planning for collaborators around the world. 2017 highlights include:

- Collaboration on an exhibition with conceptual artist Joseph Rossano, *Conservation from Here,* at the Oyster Bay Historical Society and Sagamore Hill National Historic Site.
- An NSF MaizeCODE project website that provides researchers access to genome data for the maize inbred lines B73 and NC350.
- Design work on the *DNA Subway* Purple Line.
- Designs for *Breakthrough Junior Challenge* labs, as well as production of a video featuring biodiversity barcoding at the Philippine Science High School.
- Video interviews of student research teams from *Barcode Long Island* and the *Urban Barcode Project*.
- Enhanced design of the *Mystery of Anastasia* interactive to take full advantage of the new, higher-resolution laptops in the Bioinformatics Laboratory.

Staff and Interns

The ranks of the instructional team were bolstered when Pauline McGlone, college intern, and Michael Paul, laboratory research technician, were promoted to genetics educators. Pauline began as a high school intern in 2012 and then returned while a student at Hunter College. In addition to teaching, Pauline manages our high school interns. Michael Paul joined the staff as our laboratory research technician in June. Michael's background in chemistry from Bowdoin College and natural teaching ability led to his transition to instructor and manager of college interns.

Jorge Pérez de Acha joined the DNALC's multimedia department as a Bioinformatics Researcher. Born and raised in Mexico City, Jorge attended a high school of fine arts in Boston, Massachusetts, with the intention of becoming an opera singer. However, a stint in his uncle's tax consulting firm piqued his interest in computer engineering, which he pursued at the Instituto Technológico Autónomo de México. Jorge is working on enhancing our bioinformatics analysis tools.

We said goodbye to four staff members in 2017: administrative manager Carolyn Reid Faughnan; data science educator Joslynn Lee; genetics educator Katie McAuley; and lab technician Emtiaz Uddin. Carolyn became part of the DNALC's administrative team in 2003, rising to organize the entire administrative operation—including academic year field trips, summer camp programs, visitation statistics, and so much more. Carolyn retired at year's end to enjoy her extended family and some well deserved rest and relaxation. "Ya'at'eeh! Shi éí Joslynn Lee yinishyé," was Joslynn's way of introducing herself in Navajo when she became the first Native American DNALC staff member. With a background in computational chemistry, she led the development of metagenomics analysis for CyVerse and *Barcode Long Island*. Joslynn left CSHL to become a science education fellow at Howard Hughes Medical Institute. Katie joined the DNALC in 2012, managing interns, overseeing footlocker preparation, and training DNALC Asia staff in Suzhou, in addition to instructing middle and high school labs. She left in September to teach forensic science, living environment, and earth science at Patchogue-Medford High School. With the expansion of our barcoding programs, Emtiaz became the DNALC's first full-time lab technician in 2015. He left early in 2017 to begin a career in mental health services administration.

Since the DNALC opened, we have relied on high school and college interns to support day-to-day operations. An internship offers each student the unique opportunity to gain real laboratory or design experience in an educational environment. The *BioMedia* Group also welcomes interns for summer or longer-term roles. We gathered an amazing group of interns this year, and said farewell as others left for college.

High School Interns

Duardo Akerle, Half Hollow Hills High School
Caroline Ambriano, Cold Spring Harbor High School
Elijah Calle, Hempstead High School
Randy Diaz Arias, New Heights Academy Charter School
Megan Erhardt, Huntington High School
Matthew Finkelberg, Syosset High School
Brianna Hines, Our Lady of Mercy Academy
George Homenides, Commack High School
Alyssa Lugo-Mercado, West Hempstead High School
Brady Lyons, St. Dominic's High School
Jillian Maturo, Syosset High School
Natalie McCann, Huntington High School
Erika Mosso, Aquinas High School
Salvatore Salerno, Bethpage High School
Jon Triscari, St. Anthony's High School
Nathaniel Wang, Northport High School

High School Interns Departing for College

Alyssa D'Arrigo, Stony Brook University
Alec Haber, Washington University in St. Louis
Derek Lee, Cornell University
Daniella Pillco, State University of New York at Buffalo
Rahul Ranjan, Stony Brook University
Ben Rhee, Harvard University
Bijia Wang, Emory University
GraceAnne Woods, Stony Brook University

College Interns

Kathryn Bellissimo, The College of New Jersey
Gabrielle Blazich, Fordham University
Juliana Eastment, University of Richmond
Marie Jean Francois, City University of New York at City College
Omotayo Ikuomenisan, Hunter College
Michaela Lee, State University of New York at Oneonta
William McBrien, Stony Brook University
Alec Micklos, Baruch College
Andrew Micklos, Lackawanna College
Stefanie Montalbano, Fairfield University
Breanna Tahany, State University of New York at Binghamton
Maria Urbina, Tufts University

Site Visits 2017

January 5 — *Urban Barcode Project/Urban Barcode Research Program* Open Lab, *Harlem DNA Lab*
January 11 — NSF CyVerse Webinar Series, "Get Started with CyVerse," DNALC
January 14 — *Urban Barcode Project/Urban Barcode Research Program* Open Lab, *Harlem DNA Lab*
January 14–18 — International Plant and Animal Genome XXV Conference 2017, CyVerse Education Session: *DNA Subway, Data Science, and Microbial Genomics;* "Data Science Challenges and Solutions for Student Microbiome Research," "Student DNA Genotyping by Mass Spectrometry: An Affordable Look into DNA Ancestry," "Planning, Executing and Cultivating Broader Impact Programs: Tools, Communities and Resources to Get Organized, Get Connected, Get Noticed and Get Funded," San Diego, California
January 17 — *Urban Barcode Project/Urban Barcode Research Program* Open Lab, *Harlem DNA Lab*
Urban Barcode Project Open Lab, Genspace, Brooklyn, New York
January 21 — NIH *Barcode Long Island*, *Microbiome Project* Open Lab, Brookhaven National Laboratory, Upton, New York
January 25 — *Ötzi the Iceman* Tour, DNALC
Site visit by Karen Scharbach, Jim Medina, Paul Paino, Don Corrao, and Clifford Packingham, St. Anthony's High School, South Huntington, New York
January 26 — Site visit by Donna Moro, Jack Abrams STEM School, Huntington, New York
January 27 — Cold Spring Harbor High School, *Marine Biology/DNA Barcoding* Teacher Workshop, DNALC
January 28–29 — NIH *Barcode Long Island*, *Microbiome Project* Bioinformatics Refresher, DNALC
January 29–30 — "Meet the Scientist," St. David's School, New York, New York
January 31 — *Urban Barcode Project* Open Lab, Genspace, Brooklyn, New York
NIH *Barcode Long Island* Open Lab, DNALC
February 2 — *Urban Barcode Project* Open Lab, *Harlem DNA Lab*
February 3, 5 — *ASM Agar Art* Teacher Workshop, Part 1, School of Visual Arts, New York, New York
February 4 — *Saturday DNA!* "Mitochondrial DNA Analysis," DNALC
Ötzi the Iceman Tour, DNALC
Urban Barcode Project/Urban Barcode Research Program Open Lab, *Harlem DNA Lab*
February 6–17 — *DNA Barcoding*, *DNA Science,* and *Research* Student Workshops, Beijing 166 School, DNALC
February 7 — CSHL *Gramene: A Resource for Comparative Plant Genomics* Webinar, DNALC
February 8 — NSF CyVerse Webinar Series, "Get Started with CyVerse," DNALC
February 10 — *Urban Barcode Research Program* Student Mixer, The Irondale Center for Theater, Education, and Outreach, Brooklyn, New York
February 13–14 — NIH Virtual *Train the Trainers* Workshop, DNALC

February 14	*Urban Barcode Project* Open Lab, *Harlem DNA Lab*
	Urban Barcode Project Open Lab, Genspace, Brooklyn, New York
February 17	NSF CyVerse site visit by Parker Antin, University of Arizona, Tucson, Arizona
February 18	*Urban Barcode Project/Urban Barcode Research Program* Open Lab, *Harlem DNA Lab*
February 21	*Ötzi the Iceman* Tour, DNALC
February 21–24	*Urban Barcode Research Program Conservation Genetics* Workshop, *Harlem DNA Lab*
	Urban Barcode Research Program Conservation Genetics Workshop, Rockefeller University, New York, New York
February 23	*Ötzi the Iceman* Tour, DNALC
	Site visit by Nurit Bar-Shai and Franklin Adams, Genspace, Brooklyn, New York
February 25	*Saturday DNA!* "Enzymes in Action," DNALC
February 28	*Urban Barcode Project/Urban Barcode Research Program* Open Lab, *Harlem DNA Lab*
	Urban Barcode Project Open Lab, Genspace, Brooklyn, New York
March 4	NIH *Barcode Long Island* Open Lab, DNALC
	Urban Barcode Project/Urban Barcode Research Program Open Lab, *Harlem DNA Lab*
March 8	NSF CyVerse Webinar Series, "Get Started with CyVerse," DNALC
March 9	*Urban Barcode Project* Open Lab, *Harlem DNA Lab*
March 10–13	Data Collection from Maize Community for MaizeCODE Project at Maize Genetic Conference, St. Louis, Missouri
March 11	*Saturday DNA!* "Driving Evolution," DNALC
March 14	*Urban Barcode Project/Urban Barcode Research Program* Open Lab, *Harlem DNA Lab*
March 17	*Ötzi the Iceman* Tour, DNALC
March 18–19	*Urban Barcode Project/Urban Barcode Research Program* Open Lab, *Harlem DNA Lab*
	NIH *Barcode Long Island* Open Lab, Brookhaven National Laboratory, Upton, New York
March 20–21	NSF CyVerse Researchers Workshop, "Overview of CyVerse," University of York, York, United Kingdom
March 23	*Urban Barcode Project* Open Lab, *Harlem DNA Lab*
	Long Island Science and Engineering Fair, Inc. Judging, Plainview, New York
March 25–26	NIH *Barcode Long Island* Open Lab, DNALC
March 27–30	*DNA Science/DNA Barcoding/Independent Research Internship* Teacher Workshop, Beijing 166 School, Beijing, China
March 28	*Urban Barcode Project* Open Lab, *Harlem DNA Lab*
March 29	*Ötzi the Iceman* Tour, DNALC
April 1	*Urban Barcode Project/Urban Barcode Research Program* Open Lab, *Harlem DNA Lab*
April 4	*Urban Barcode Research Program* Open Lab, *Harlem DNA Lab*
April 5	NSF CyVerse Webinar Series, "Get Started with CyVerse," DNALC
April 6	*Urban Barcode Project/Urban Barcode Research Program* Open Lab, *Harlem DNA Lab*
April 11	CSHL *Gramene: A Resource for Comparative Plant Genomics* Webinar, DNALC
April 13–14	Big Data Literacy Workshop, New York Hall of Science, Queens, New York
April 18	2017 LI STEM Hub Annual Celebration and Industry & Student Showcases, Brookhaven National Laboratory, Upton, New York
April 18–19	AgBioData Meeting, "CyVerse Overview," Salt Lake City, Utah
April 20	*Urban Barcode Project* Open Lab, *Harlem DNA Lab*
	"Native American Role Model Speaker Series," Northern Arizona University, Flagstaff, Arizona
	Site visit by Raymond Loverso and Syosset High School science faculty members, DNALC
April 20–21	Site visit by Mexican Delegation: Mauricio Hernandez, former Director National Public Health Institute and former Vice Minister of Health; Liliana Hernandez, IAP, Mexico; Armando Barriguete, Coordinator, DNALC Mexico
April 21	*Barcode Long Island* at Suffolk Teachers Association of New York State
	Stem Conference, Brookhaven National Laboratory, Upton, New York
April 22	*Urban Barcode Project/Urban Barcode Research Program* Open Lab, *Harlem DNA Lab*
	Saturday DNA! "DNA on the Go," DNALC
	Ötzi the Iceman Tour, DNALC
	NIH *Barcode Long Island* Open Lab, Brookhaven National Laboratory, Upton, New York
	Science March for Earth Day, New York, New York
April 26	*Ötzi the Iceman* Tour, DNALC
April 29–30	NIH *Barcode Long Island* Open Lab, DNALC
May 2	*Urban Barcode Project* Open Lab, *Harlem DNA Lab*
May 4	*DNA Subway* Open Lab, DNALC
	Ötzi the Iceman Tour, DNALC

May 6	*Urban Barcode Research Program* Open Lab, *Harlem DNA Lab*
May 6–7	NIH *Barcode Long Island* Open Lab, Stony Brook University, Stony Brook, New York
May 9	*Urban Barcode Project*/*Urban Barcode Research Program* Open Lab, *Harlem DNA Lab*
May 10	NSF CyVerse Webinar Series, "Get Started with CyVerse," DNALC
	Urban Barcode Research Program Student Mixer, The Irondale Center for Theater, Education, and Outreach, Brooklyn, New York
May 12	CUNY Advanced Science Research Center Urban Collaborative Day, New York, New York
	Site visit by Julie Horvath, North Carolina Museum of Natural Sciences, Raleigh, North Carolina
May 13	*Saturday DNA!* "Be WiSE about Neuroscience," DNALC
	NIH *Barcode Long Island* Open Lab, Roslyn High School, Roslyn, New York
	DNA Subway Open Lab, DNALC
May 16	NIH *Barcode Long Island,* Shoreham-Wading River High School STEM Symposium, Wading River, New York
May 15–17	NIH Microbiome Hackathon, DNALC
May 20	NIH *Barcode Long Island* Open Lab, Brookhaven National Laboratory, Upton, New York
	Banana DNA Extraction Exhibition, Super Saturday STEM Expo, Harlem Children's Zone Armory, New York, New York
May 22–26	NSF CyVerse *Genome Science and Leadership* Workshop, Durango, Colorado
May 24	*Ötzi the Iceman* Tour, DNALC
May 25	*Urban Barcode Project*/Pinkerton *Urban Barcode Research Program* Symposium, New York Academy of Medicine, New York, New York
May 30	*Urban Barcode Research Program* at High School for Environmental Studies Science Fair, New York, New York
May 31	NIH *Barcode Long Island*, 2017 Open Space Stewardship Celebration/Symposium, Brookhaven National Laboratory, Upton, New York
	Urban Barcode Project at Tenafly Science Fair, Tenafly, New Jersey
June 2	"Approaches to Evaluating Authentic Science Research," NIH SciEd Conference 2107, Metro Center, Washington, D.C.
June 3	*Saturday DNA!* "The Extraction Attraction," DNALC
June 6	"DNA Barcoding," Genspace, Brooklyn, New York
	NSF CyVerse *Tools and Services* Workshop, University of Arkansas, Fayetteville, Arkansas
June 7	*Barcode Long Island* Student Symposium, CSHL
	NSF CyVerse Webinar Series, "Get Started with CyVerse," DNALC
June 7–8	NSF CyVerse *Genomics Data Carpentry* Workshop, University of Arkansas, Fayetteville, Arkansas
	Science Research Mentoring Program Colloquium with *UBRP* students, American Museum of Natural History, New York, New York
June 9	*Urban Barcode Project*/*Urban Barcode Research Program* at Youth Educational Seining Symposium, Brooklyn Bridge Park and St. Francis College, Brooklyn, New York
June 10	Cold Spring Harbor Laboratory Open House, CSHL
June 10–11	NSF CyVerse *Genomics in Education* Workshop, James Madison University, Harrisonburg, Virginia
June 12	NSF CyVerse *Tools and Services* Workshop, James Madison University, Harrisonburg, Virginia
June 12–16	*Forensic Detectives* Workshop, St. David's School, New York, New York
	Green Genes Workshop, Lycée Français, New York, New York
	Human Genomics Workshop, Lycée Français, New York, New York
June 15–16	ECSITE Conference, "Having Fun With Classification," "Collective Therapy Parody," and "Unleashing Citizen Contributions to Science for All," European Citizen Science Association Working Group, Interdisciplinary Centre of Marine and Environmental Research, Natural History and Science Museum of the University of Porto, Porto, Portugal
June 14	*Ötzi the Iceman* Tour, DNALC
June 15–26	*DNA Barcoding* Workshop, Grace Church High School, New York, New York
June 16	Site visit by Bruce Ratner, Forest City New York, Brooklyn, New York
June 19	Site visit by Project Lead Sam Janis of Billion Oysters Project, *Harlem DNA Lab*
June 20	*Urban Barcode Research Program*, Pinkerton Youth Council, American Museum of Natural History, New York, New York
June 22	Exhibited at the Long Island Invasive Species Management Area Conference, Sisters of St. Joseph, Brentwood, New York
	Ötzi the Iceman Tour, DNALC
June 23	Site visit by Mike and Lois O'Brien and Ebby Gerry, *Mary Harriman and Eugenics*, DNALC
June 26	NSF CyVerse, Plant Biology 2017, Honolulu, Hawaii
June 26–30	NIH *Barcode Long Island* Workshop, Brookhaven National Laboratory, Upton, New York
	Genome Science Workshop, DNALC Asia, Suzhou, China

	DNA Science Workshop, DNALC (2 sessions)
	Fun with DNA Workshop, DNALC
	Green Genes Workshop, DNALC
	Fun with DNA Workshop, DNA Learning Center *West*
June 27–July 8	*Genome Science* Workshop, Grace Church High School, New York, New York
July 3–7	*Forensic Detectives* Workshop, DNALC
	Fun with DNA Workshop, DNALC
	Genome Science Workshop, DNALC
	World of Enzymes Workshop, DNALC
	World of Enzymes Workshop, DNA Learning Center *West*
July 3–7	*Fun with DNA* Workshop, DNALC Asia, Suzhou, China
July 6	NIH CyVerse *Tools and Services* Workshop, University of Arkansas, Fayetteville, Arkansas
July 7	*Alu* Lab, Study @ Dushu Lake Tour from Suzhou, China, DNALC
July 7–8	NIH CyVerse *Genomics Data Carpentry* Workshop, University of Arkansas, Fayetteville, Arkansas
July 10	*Ötzi the Iceman* Tour, DNALC
	Alu Lab, Suzhou, China group, DNALC
July 10–14	*Barcode Long Island* Workshop, DNALC
	DNA Science Workshop, DNALC
	Green Genes Workshop, DNALC
	World of Enzymes Workshop, DNALC
	DNA Science Workshop, DNA Learning Center *West*
	Pinkerton *Urban Barcode Research Program Conservation Genetics* Workshop, *Harlem DNA Lab*
	World of Enzymes Workshop, DNALC Asia, Suzhou, China
July 13	"Barcode Long Island: Infrastructure for Citizen Science," DIYbio—DC Genomics Meetup, NIH, Bethesda, Maryland
July 14	"DNA Barcoding: Infrastructure for Citizen Science," Citizen Science Working Group, NIH, Bethesda, Maryland
July 17–21	Pinkerton *Urban Barcode Research Program DNA Barcoding* Workshop, *Harlem DNA Lab*
	Fun with DNA/World of Enzymes Workshops, Beijing 166 School, DNALC
	DNA Barcoding Workshop, DNALC
	Fun with DNA Workshop, DNALC
	Green Genes Workshop, DNA Learning Center *West*
July 19	NSF CyVerse Webinar Series, "Get Started with CyVerse," DNALC
July 24	*Ötzi the Iceman* Tour, DNALC
July 24–28	*Forensics* Workshop, Beijing 166 School, DNALC
	Green Genes Workshop, Beijing 166 School, DNALC
	BioCoding Workshop, DNALC
	DNA Science Workshop, DNALC
	World of Enzymes Workshop, DNALC
	Fun with DNA Workshop, DNA Learning Center *West*
July 25	*Ötzi the Iceman* Tour, DNALC
	Professional Development Workshop, "Bacterial Transformation and Protein Purification," *Harlem DNA Lab*
July 26	Professional Development Workshop, "Restriction Analysis," *Harlem DNA Lab*
July 27	Professional Development Workshop, "DNA Fingerprinting—Detecting *Alu* Mutation by PCR," *Harlem DNA Lab*
July 28	Professional Development Workshop, "Human Mitochondrial Sequencing," *Harlem DNA Lab*
	Presentation, Shanghai Business School and Shanghai Finance University, CSHL
July 31	*Ötzi the Iceman* Tour, DNALC
July 31–Aug 4	*Forensics* Workshop, Beijing 166 School, DNALC
	Green Genes Workshop, Beijing 166 School, DNALC
	DNA Barcoding Workshop, DNALC
	Fun with DNA Workshop, DNALC
	DNA Science Workshop, DNA Learning Center *West*
	DNA Science Workshop, *Harlem DNA Lab*
	Fun with DNA Workshop, New World Preparatory Charter School, Staten Island, New York
August 7	*Ötzi the Iceman* Tour, DNALC
August 7–11	*DNA Science* Workshop, DNALC
	Fun with DNA Workshop, DNALC
	Green Genes Workshop, DNALC
	World of Enzymes Workshop, DNALC

	Forensic Detectives Workshops, DNA Learning Center *West*
	Pinkerton *Urban Barcode Research Program Conservation Genetics* Workshop, *Harlem DNA Lab*
	Barcode Long Island Workshop, Stony Brook University, Stony Brook, New York
August 9	NSF CyVerse Webinar Series, "Get Started with CyVerse," DNALC
August 14	*Ötzi the Iceman* Tour, DNALC
	Site visit by Jenny Negrom, Pinkerton Foundation, *Harlem DNA Lab*
August 14–18	*DNA Science* Workshop, DNALC
	Fun with DNA Workshop, DNALC
	Genome Science Workshop, DNALC
	Green Genes Workshop, DNALC
	Green Genes Workshop, DNA Learning Center *West*
	World of Enzymes Workshop, New World Preparatory Charter School, Staten Island, New York
	Pinkerton *Urban Barcode Research Program DNA Barcoding* Workshop, *Harlem DNA Lab*
August 15	Site visit by Carlos Sierra Sanchez, Columbia University, *Harlem DNA Lab*
August 16	Site visit by Guangzhou Delegation, including Zhou Yawei, standing member of the CPC Guangzhou Committee and Director-General of Guangzhou Development District Administrative Committee, and Sherry Liu, Guangzhou, China
August 21	Site visit by Martin Elias of Elias Properties and family
August 21–25	*Being Human* Workshop, DNALC
	DNA Science Workshop, DNALC
	Forensic Detectives Workshop, DNALC
	World of Enzymes Workshop, DNALC
	Fun with DNA Workshop, DNA Learning Center *West*
	DNA Barcoding Teacher Workshop, *Harlem DNA Lab*
August 22	Eugenics Interview, *White Supremacy, Eugenics, and the History of Selective Breeding*, NBC Left Field, CSHL
August 24–25	*Software Carpentry* Workshop, Los Alamos National Lab, New Mexico
August 28–29	NSF CyVerse *Genomics Data Carpentry* Workshop, New Mexico State University, Las Cruces, New Mexico
Aug 28–Sept 1	*DNA Science* Workshop, DNALC
	Fun with DNA Workshop, DNALC
	Green Genes Workshop, DNALC
	World of Enzymes Workshop, DNALC
	WiSE *Fun with DNA* Workshop, CSHL
	World of Enzymes Workshop, DNA Learning Center *West*
August 30	NSF CyVerse *Data Carpentry* Workshop, New Mexico State University, Las Cruces, New Mexico
September 11	Site visit by Hannah Meagher and Mark Mannucci, American Masters series, to film Dr. James Watson, DNALC
September 13	NSF CyVerse Webinar Series, "Get Started with CyVerse," DNALC
September 16	Pine Barrens Discovery Day Event, Wertheim Wildlife Refuge, Shirley, New York
September 19	*Ötzi the Iceman* Tour, DNALC
	Site visit by Brett Curlew, BNL/BOCES and OceansWide, DNALC
September 21	*Ötzi the Iceman* Tour, DNALC
October 3	*Urban Barcode Research Program* Cohort 5 Mentor Matching Event, The Irondale Center for Theater, Education, and Outreach, Brooklyn, New York
October 6	Site visit by Jeff Chen, Suzhou High School of Jiangsu Province, Suzhou, China
	"The *Urban Barcode Research Program*," CUNY City College, New York, New York
October 14	*Saturday DNA!* "Dust Away Crime—The Truth About Fingerprints," DNALC
October 15	*Ötzi the Iceman* Tour, DNALC
October 16–17	*Human DNA Fingerprint: Genotyping a "Jumping Gene"* Teacher Workshop, Math for America, New York, New York
	PTC/GMO Workshop, Beijing 166 School, Beijing, China
October 18	NSF CyVerse Webinar Series, "Get Started with CyVerse," DNALC
October 19	*Ötzi the Iceman* Workshop, Michael J. Petrides School, Staten Island, New York
October 18, 20	PTC/GMO Workshop, Beijing 166 School, Beijing, China
October 21	Sagamore Hill STEM Festival Event, *DNA Extractions*, Oyster Bay, New York
October 23	Site visit by Jeffry Petracca, Long Island Aquarium, Riverhead, New York
October 23–24	*PTC/GMO* Workshop, Beijing 166 School, Beijing, China
October 24	Site visit by Beverly Lee-Wo, The Nature's Bounty Company, Ronkonkoma, New York
October 26	*Urban Barcode Research Program*, Austrian Research and Innovation Talks (ARIT), Palm Door on Sixth, Austin, Texas
	DNA Barcoding Teacher Workshop, Beijing 166 School, Beijing, China

November 2	*DNA Barcoding* Teacher Workshop, Beijing 166 School, Beijing, China
November 3	Site visit by Suzhou Industrial Park Administrative Committee Education Bureau Delegation, Suzhou, China
November 4	*Saturday DNA!* "Glowing Genes," DNALC
November 10–11	*Ötzi the Iceman* Tour, DNALC
	"Forensic Analysis of Ötzi the Iceman," "Sense in Molecules: Modeling Personalized Medicine," and "DNA Barcoding—Independent Research in the Classroom," NABT—2017 Professional Development Conference, St. Louis, Missouri
November 13–17	*Breakthrough Junior Challenge* Teacher Training Workshop, Palawan, Philippines
November 14	Site visit by Allison Slabaugh, Mary Galvin, and Amy Stark, Notre Dame University, South Bend, Indiana
November 15	NSF CyVerse Webinar Series, "Get Started with CyVerse," DNALC
November 18	NIH *Barcode Long Island* Open Lab, DNALC
November 21	*Ötzi the Iceman* Retirees Tour and Film: "The Transition Network," DNALC
November 29	"The Bioinformatics Training Landscape," Toronto Area Bioinformatics User Group (TorBUG), University of Toronto, Toronto, Canada
December 2	NIH *Barcode Long Island* Open Lab, DNALC
December 9	*Saturday DNA!* "A Royal Ruse," DNALC
December 13	NSF CyVerse Webinar Series, "Get Started with CyVerse," DNALC
December 16	NIH *Barcode Long Island* Open Lab, Stony Brook University, Stony Brook, New York
December 21	Site visit by Dr. Armando Mansilla Olivares and family, Mexican National Academy of Medicine and Armando Barriguete, Mexico City, Mexico
December 27	*Ötzi the Iceman* Tour, DNALC
December 28	*Ötzi the Iceman* Tour, DNALC

Sites of Major Faculty Workshops

Program Key: *Middle School* High School **College**

State	*Institution*	*Year(s)*
ALABAMA	University of Alabama, Tuscaloosa	1987–1990
	Hudson Alpha Institute, Huntsville	**2014**
ALASKA	**University of Alaska, Anchorage**	**2012**
	University of Alaska, Fairbanks	**1996**
ARIZONA	Arizona State University, Tempe	2009
	Tuba City High School	1988
	University of Arizona, Tucson	**2011**
	United States Department of Agriculture, Maricopa	**2012**
ARKANSAS	Henderson State University, Arkadelphia	1992
	University of Arkansas, Fayetteville	**2017**
	University of Arkansas, Little Rock	**2012**
CALIFORNIA	**California State University, Dominguez Hills**	**2009**
	California State University, Fullerton	**2000**
	California State University, Long Beach	**2015**
	California Institute of Technology, Pasadena	**2007**
	Canada College, Redwood City	1997
	City College of San Francisco	2006
	City College of San Francisco	**2011, 2013**
	Community College of Denver	**2014**
	Contra Costa County Office of Education, Pleasant Hill	2002, 2009
	Foothill College, Los Altos Hills	**1997**
	Harbor-UCLA Research & Education Institute, Torrance	2003
	Los Angeles Biomedical Research Institute (LA Biomed), Torrance	**2006**
	Laney College, Oakland	1999
	Lutheran University, Thousand Oaks	1999
	Oxnard Community College, Oxnard	2009
	Pasadena City College	**2010**
	Pierce College, Los Angeles	**1998**
	Salk Institute for Biological Studies, La Jolla	2001, 2008
	San Francisco State University	**1991**
	San Diego State University	**2012**

	San Jose State University	**2005**
	Santa Clara University	2010
	Southwestern College, Chula Vista	**2014–2015**
	Stanford University, Palo Alto	**2012**
	University of California, Berkeley	**2010, 2012**
	University of California, Davis	1986
	University of California, Davis	**2012, 2014–2015**
	University of California, Long Beach	**2015**
	University of California, Northridge	**1993**
	University of California, Riverside	2011
	University of California, Riverside	**2012**
	University of California, San Francisco	**2015**
COLORADO	Aspen Science Center	2006
	Colorado College, Colorado Springs	1994, 2007
	Colorado State University, Fort Collins	**2013**
	Community College of Denver	**2014**
	United States Air Force Academy, Colorado Springs	**1995**
	University of Colorado, Denver	1998, 2009–2010
CONNECTICUT	Choate Rosemary Hall, Wallingford	1987
DELAWARE	**Jackson Laboratory, Farmington University of Delaware, Newark**	**2016**
DISTRICT OF COLUMBIA	**Howard University, Washington**	**1992, 1996, 2009–2010**
FLORIDA	Armwood Senior High School, Tampa	1991
	Florida Agricultural & Mechanical University, Tallahassee	2007–2008
	Florida Agricultural & Mechanical University, Tallahassee	**2011**
	Florida SouthWestern State University, Fort Myers	**2015**
	North Miami Beach Senior High School	1991
	Seminole State College, Sanford	**2013–2014**
	University of Miami School of Medicine	2000
	University of Western Florida, Pensacola	1991
GEORGIA	Fernbank Science Center, Atlanta	1989, 2007
	Gwinnett Technical College, Lawrenceville	**2011–2012**
	Morehouse College, Atlanta	**1991, 1996–1997**
	Spelman College, Atlanta	**2010**
	University of Georgia, Athens	**2015**
HAWAII	Kamehameha Secondary School, Honolulu	1990
	University of Hawaii at Manoa	**2012**
IDAHO	University of Idaho, Moscow	1994
ILLINOIS	Argonne National Laboratory	1986–1987
	iBIO Institute/Harold Washington College, Chicago	2010
	Illinois Institute of Technology, Chicago	2009
	Kings College, Chicago	**2014**
	University of Chicago	**1992, 1997, 2010**
	University of Southern Illinois, Carbondale	**2016**
INDIANA	Butler University, Indianapolis	1987
	Purdue University, West Lafayette	**2012**
IOWA	Drake University, Des Moines	1987
KANSAS	University of Kansas, Lawrence	1995
KENTUCKY	**Bluegrass Community & Technical College, Lexington**	**2012–2014**
	Murray State University	1988
	University of Kentucky, Lexington	1992
	Western Kentucky University, Bowling Green	1992
LOUISIANA	Bossier Parish Community College	2009
	Jefferson Parish Public Schools, Harvey	1990
	John McDonogh High School, New Orleans	1993
	Southern University at New Orleans	**2012**
MAINE	Bates College, Lewiston	1995
	Southern Maine Community College	**2012–2013**
	Foundation for Blood Research, Scarborough	2002
MARYLAND	Annapolis Senior High School	1989

	Bowie State University	**2011, 2015**
	Frederick Cancer Research Center	1995
	McDonogh School, Baltimore	1988
	Montgomery County Public Schools	1990–1992
	National Center for Biotechnology Information, Bethesda	2002
	St. John's College, Annapolis	*1991*
	University of Maryland, School of Medicine, Baltimore	**1999**
MASSACHUSETTS	**Arnold Arboretum of Harvard University, Roslindale**	**2011**
	Beverly High School	1986
	Biogen Idec, Cambridge	2002, 2010
	Boston University	**1994, 1996**
	CityLab, Boston University School of Medicine	1997
	Dover-Sherborn High School, Dover	1989
	Randolph High School	1988
	The Winsor School, Boston	1987
	Whitehead Institute for Biomedical Research, Cambridge	2002
MICHIGAN	Athens High School, Troy	1989
	Schoolcraft College, Livonia	**2012**
MINNESOTA	**American Society of Plant Biologists, Minneapolis**	**2015**
	Minneapolis Community and Technical College, Madison	2009
	Minneapolis Community and Technical College, Madison	**2013**
	University of Minnesota, St. Paul	2005
	University of Minnesota, St. Paul	**2010**
MISSISSIPPI	Mississippi School for Math & Science, Columbus	1990–1991
	Rust College, Holly Springs	**2006–2008, 2010**
MISSOURI	**St. Louis Science Center**	**2008–2010**
	Stowers Institute for Medical Research, Kansas City	2002, 2008
	University of Missouri, Columbia	**2012**
	Washington University, St. Louis	**1989, 1997, 2011**
MONTANA	**Montana State University, Bozeman**	**2012**
NEBRASKA	**University of Nebraska-Lincoln, Lincoln**	**2014**
NEVADA	University of Nevada, Reno	1992, **2014**
NEW HAMPSHIRE	Great Bay Community College, Portsmouth	2009
	New Hampshire Community Technical College, Portsmouth	**1999**
	St. Paul's School, Concord	1986–1987
NEW JERSEY	Coriell Institute for Medical Research, Camden	2003
	Raritan Valley Community College, Somerville	2009
NEW MEXICO	Biolink Southwest Regional Meeting, Albuquerque	2008
	Los Alamos National Lab	**2017**
	New Mexico State University, Las Cruces	**2017**
	Santa Fe Community College, Santa Fe	**2015**
NEW YORK	Albany High School	1987
	American Museum of Natural History, New York	2007, 2015
	Bronx High School of Science	1987
	Brookhaven National Laboratory, Upton	2015–2017
	Canisius College, Buffalo	2007
	Canisius College, Buffalo	**2011**
	City College of New York	2012
	Cold Spring Harbor High School	1985, 1987
	Cold Spring Harbor Laboratory	**2014–2015**
	Columbia University, New York	**1993**
	Cornell University, Ithaca	**2005**
	DeWitt Middle School, Ithaca	*1991, 1993*
	Dolan DNA Learning Center	1988–1995, 2001–2004, 2006–2009, 2015–2017
	Dolan DNA Learning Center	**1990, 1992, 1995, 2000–2011**
	Dolan DNA Learning Center	*1990–1992*
	DNA Learning Center *West*	2005
	Environmental Science Center, Bergen Beach, Brooklyn	2015–2016
	Fostertown School, Newburgh	*1991*

	Harlem DNA Lab, East Harlem	2008–2009, 2011–2013, 2016, 2017
	***Harlem DNA Lab*, East Harlem**	**2015–2016**
	Huntington High School	1986
	Irvington High School	1986
	John Jay College of Criminal Justice	2009
	Junior High School 263, Brooklyn	*1991*
	Lindenhurst Junior High School	*1991*
	Math for America	**2017**
	Mount Sinai School of Medicine, New York	1997
	Nassau Community College, Garden City	**2013**
	New York Botanical Garden, Bronx	**2013**
	New York City Department of Education	2007, 2012
	New York Institute of Technology, New York	2006
	New York Institute of Technology, New York	**2006**
	Orchard Park Junior High School	*1991*
	Plainview-Old Bethpage Middle School	*1991*
	School for Visual Arts, New York	**2017**
	State University of New York, Purchase	1989
	State University of New York, Stony Brook	1987–1990, 2015–2017
	State University of New York, Stony Brook	**2014, 2016**
	Stuyvesant High School, New York	1998–1999
	The Rockefeller University, New York	2003, 2015–2016
	The Rockefeller University, New York	**2010**
	Titusville Middle School, Poughkeepsie	*1991, 1993*
	Trudeau Institute, Saranac Lake	2001
	Union College, Schenectady	2004
	United States Military Academy, West Point	**1996**
	Wheatley School, Old Westbury	1985
NORTH CAROLINA	CIIT Center for Health Research, Triangle Park	2003
	North Carolina Agricultural & Technical State University, Greensboro	**2006–2007, 2009–2011**
	North Carolina School of Science, Durham	1987
	North Carolina State University, Raleigh	**2012**
NORTH DAKOTA	**North Dakota State University, Fargo**	**2012**
OHIO	Case Western Reserve University, Cleveland	1990
	Cleveland Clinic	1987
	Langston University, Langston	2008
	North Westerville High School	1990
	The Ohio State University, Wooster	**2016**
OKLAHOMA	**Oklahoma City Community College**	**2000**
	Oklahoma City Community College	2006–2007, 2010
	Oklahoma Medical Research Foundation, Oklahoma City	2001
	Oklahoma School of Science and Math, Oklahoma City	1994
	Tulsa Community College, Tulsa	2009
	Tulsa Community College, Tulsa	**2012–2014**
OREGON	Kaiser Permanente-Center for Health Research, Portland	2003
	Linfield College, McMinnville	**2014**
PENNSYLVANIA	Duquesne University, Pittsburgh	1988
	Germantown Academy	1988
	Kimmel Cancer Center, Philadelphia	2008
RHODE ISLAND	Botanical Society of America, Providence	2010
SOUTH CAROLINA	**Clemson University**	**2004, 2015**
	Medical University of South Carolina, Charleston	1988
	University of South Carolina, Columbia	**1988**
SOUTH DAKOTA	**South Dakota State University, Brookings**	**2015**
TENNESSEE	NABT Professional Development Conference, Memphis	2008
TEXAS	Austin Community College–Rio Grande Campus	2000
	Austin Community College–Eastview Campus–Roundrock Campus	**2007–2009, 2013**

	Austin Community College–Roundrock Campus	**2012**
	Houston Community College Northwest	**2009–2010**
	J.J. Pearce High School, Richardson	1990
	Langham Creek High School, Houston	1991
	University of Lone Star College, Kingwood	**2011**
	Midland College	2008
	Southwest Foundation for Biomedical Research, San Antonio	2002
	Taft High School, San Antonio	1991
	Texas A&M University, College Station, TX	**2013**
	Texas A&M University, Prairie View, TX	**2013**
	Texas A&M, AG Research and Extension Center, Weslaco	2007
	Trinity University, San Antonio	**1994**
	University of Texas, Austin	**1999, 2004, 2010, 2012**
	University of Texas, Brownsville	**2010**
UTAH	**Brigham Young University, Provo**	**2012**
	University of Utah, Salt Lake City	1993
	University of Utah, Salt Lake City	**1998, 2000**
	Utah Valley State College, Orem	**2007**
VERMONT	University of Vermont, Burlington	1989
VIRGINIA	Eastern Mennonite University, Harrisonburg	1996
	James Madison University, Harrisonburg	**2017**
	Jefferson School of Science, Alexandria	1987
	Mathematics and Science Center, Richmond	1990
	Mills Godwin Specialty Center, Richmond	1998
	Virginia Polytechnic Institute and State University, Blacksburg	**2005, 2008–2009**
WASHINGTON	Fred Hutchinson Cancer Research Center, Seattle	1999, 2001, 2008
	Shoreline Community College	2011, 2012
	University of Washington, Seattle	**1993, 1998, 2010**
WEST VIRGINIA	Bethany College	1989
WISCONSIN	Blood Center of Southeastern Wisconsin, Milwaukee	2003
	Madison Area Technical College/Madison Area College	**1999, 2009, 2011–2014**
	Marquette University, Milwaukee	1986–1987
	University of Wisconsin, Madison	1988–1989
	University of Wisconsin, Madison	**2004, 2012**
WYOMING	University of Wyoming, Laramie	1991
PUERTO RICO	**Universidad del Turabo, Gurabo, Puerto Rico**	**2011, 2012, 2014**
	University of Puerto Rico, Mayaguez	1992
	University of Puerto Rico, Mayaguez	**1992**
	University of Puerto Rico, Rio Piedras	**1993**
	University of Puerto Rico, Rio Piedras	1994

AUSTRALIA	Walter and Eliza Hall Institute and University of Melbourne	1996
	EMBL/Australian Bioinformatics Resource, University of Melbourne	**2016**
AUSTRIA	**Vienna Open Lab, Vienna**	**2007, 2012**
CANADA	Red River Community College, Winnipeg, Manitoba	1989
CHINA	Beijing No. 166 High School, Beijing	2013, 2014–2017
	Ho Yu College, Hong Kong	**2009**
DENMARK	**Faroe Genome Project, Torshavn, Faroe Islands**	**2013**
GERMANY	Urania Science Center, Berlin	2008
IRELAND	**European Conference on Computational Biology/Intelligent System for Molecular Biology Conference, Dublin**	**2015**
ITALY	International Institute of Genetics and Biophysics, Naples	1996
	Porto Conte Research and Training Laboratories, Alghero	1993
MEXICO	ADN Mexico, Morelia	2016
	ASPB Plant Biology, Merida	2008
	Langebio/Cinvestav, Irapuato	**2016**
NIGERIA	Godfrye Okoye University, Enugu, Nigeria	2013
PHILIPPINES	Eastern Visayas Campus, Philippine Science High School, Palo, Leyte	2017

RUSSIA	Shemyakin Institute of Bioorganic Chemistry, Moscow	1991
SINGAPORE	National Institute of Education	2001–2005
	Singapore Science Center	2013
SOUTH AFRICA	**North-West University, Potchefstroom**	**2016**
	South African Bioinformatics Society, Durban	**2016**
SWEDEN	Kristineberg Marine Research Station, Fiskebäckskil	1995
	Uppsala University	2000
THE NETHERLANDS	**International Chromosome Conference, Amsterdam**	**2007**
	Wageningen University and Research Center, Wageningen	**2014**
UNITED KINGDOM	**The Genome Analysis Centre, Norwich**	**2015**
	University of York, York	**2017**
	Wellcome Trust Conference Centre, Hinxton	**2012–2013**
	University of Warwick, Coventry	**2013**

COLD SPRING HARBOR
LABORATORY PRESS

PRESS PUBLICATIONS

Serials

Genes & Development, Vol. 31 (www.genesdev.org)

Genome Research, Vol. 27 (www.genome.org)

Learning & Memory, Vol. 24 (www.learnmem.org)

RNA, Vol. 23 (www.rnajournal.org)

Cold Spring Harbor Symposia in Quantitative Biology, Vol. 81: *Targeting Cancer*, edited by David Stewart and Bruce Stillman

Cold Spring Harbor Protocols, Vol. 2017 (www.cshprotocols.org)

Cold Spring Harbor Perspectives in Biology, Vol. 9 (www.cshperspectives.org)

Cold Spring Harbor Perspectives in Medicine, Vol. 7 (www.perspectivesinmedicine.org)

Cold Spring Harbor Molecular Case Studies, Vol. 3 (www.molecularcasestudies.org)

Laboratory Manuals

Ion Channels: A Laboratory Manual, edited by Paul J. Kammermeier, Ian Duguid, and Stephan Brenowitz

Textbooks

Essentials of Glycobiology, Third Edition, edited by Ajit Varki, Richard D. Cummings, Jeffrey D. Esko, Pamela Stanley, Gerald W. Hart, Markus Aebi, Alan G. Darvill, Taroh Kinoshita, Nicolle H. Packer, James H. Prestegard, Ronald L. Schnaar, and Peter H. Seeberger

Monographs (Topic Collections from *Perspectives in Biology* and *Perspectives in Medicine*)

Malaria: Biology in the Era of Eradication, edited by Dyann F. Wirth and Pedro L. Alonso

Cell Polarity, edited by Keith E. Mostov

Chromatin Deregulation in Cancer, edited by Scott A. Armstrong, Steven Henikoff, and Christopher R. Vakoc

Synthetic Biology: Tools for Engineering Biological Systems, edited by Daniel G. Gibson, Clyde A. Hutchison III, Hamilton O. Smith, and J. Craig Venter

Tissue Engineering and Regenerative Medicine, edited by Joseph P. Vacanti

Prion Diseases, edited by Stanley B. Prusiner

The Biology of Exercise, edited by Juleen R. Zierath, Michael J. Joyner, and John A. Hawley

The Biology of the TGF-β Family, edited by Rik Derynck and Kohei Miyazono

Prion Biology, edited by Stanley B. Prusiner

Cancer Evolution, edited by Charles Swanton, Alberto Bardelli, Kornelia Polyak, Sohrab P. Shah, and Trevor A. Graham

Cell–Cell Junctions, Second Edition, edited by Carien M. Niessen and Alpha S. Yap

Other

A Cure Within: Scientists Unleashing the Immune System to Kill Cancer, by Neil Canavan

CSHL Annual Report 2015, Yearbook Edition

Banbury Center Annual Report 2016

E-books

Signal Transduction: Principles, Pathways, and Processes, edited by Lewis C. Cantley, Tony Hunter, Richard Sever, and Jeremy Thorne

Epigenetics, Second Edition, edited by C. David Allis, Marie-Laure Caparros, Thomas Jenuwein, and Danny Reinberg; Associate Editor Monika Lachner

Ion Channels: A Laboratory Manual, edited by Paul J. Kammermeier, Ian Duguid, and Stephan Brenowitz

Synthetic Biology: Tools for Engineering Biological Systems, edited by Daniel G. Gibson, Clyde A. Hutchison III, Hamilton O. Smith, and J. Craig Venter

Tissue Engineering and Regenerative Medicine, edited by Joseph P. Vacanti

Chromatin Deregulation in Cancer, edited by Scott A. Armstrong, Steven Henikoff, and Christopher R. Vakoc

Cell Polarity, edited by Keith E. Mostov

Malaria: Biology in the Era of Eradication, edited by Dyann F. Wirth and Pedro L. Alonso

Cold Spring Harbor Symposia in Quantitative Biology, Vol. 81: *Targeting Cancer*, edited by David Stewart and Bruce Stillman

Prion Diseases, edited by Stanley B. Prusiner

The Biology of Exercise, edited by Juleen R. Zierath, Michael J. Joyner, and John A. Hawley

The Biology of the TGF-β Family, edited by Rik Derynck and Kohei Miyazono

Cancer Evolution, edited by Charles Swanton, Alberto Bardelli, Kornelia Polyak, Sohrab P. Shah, and Trevor A. Graham

Prion Biology, edited by Stanley B. Prusiner

A Cure Within: Scientists Unleashing the Immune System to Kill Cancer, by Neil Canavan

Cell–Cell Junctions, Second Edition, edited by Carien M. Niessen and Alpha S. Yap

Circadian Rhythms, edited by Paolo Sassone-Corsi, Michael W. Young, and Akhilesh B. Reddy

Immune Memory and Vaccines: Great Debates, edited by Shane Crotty and Rafi Ahmed

Websites

Cold Spring Harbor Monographs Archive Online (www.cshmonographs.org)

Cold Spring Harbor Symposium on Quantitative Biology Archive (symposium.cshlp.org)

Services

BioSupplyNet, scientific supply directory (www.biosupplynet.com)

COLD SPRING HARBOR LABORATORY PRESS EXECUTIVE DIRECTOR'S REPORT

Science is a global enterprise that depends on the timely communication of ideas and results. Cold Spring Harbor Laboratory Press provides scientists worldwide with authoritative, affordable, and pertinent information to further their research and aid in their career development. This ambition continues the Laboratory's commitment to scientific communication that began with the first Annual Symposium in 1933.

The Press publishes nine journals and more than 200 books in print and electronic form. The long-established journals *Genome Research* and *Genes & Development* remain preeminent in their fields, with in-house editorial teams adept in the assessment of new ideas and technologies. *RNA* and *Learning & Memory* serve specialized research communities in valuable ways. The newer review journals, *CSH Perspectives in Biology*, *CSH Perspectives in Medicine*, and *CSH Protocols*, continue to advance in stature and financial success. Each transforms content that in former years appeared only in print books, rendering it in a digital, readily discoverable, and reusable serial form. *CSH Molecular Case Studies* reimagines the traditional case report in medicine to enable open sharing of insights that genomic and molecular analysis bring to the understanding and potential treatment of disease. In its second full year, the journal had a steady increase in submissions. The Press journals overall had a record yearly download of more than 18 million full-text articles, including more than 5.5 million via PubMedCentral at the National Library of Medicine.

The newest journal is *Life Science Alliance* (*LSA*), a new open-access journal owned and published jointly by Cold Spring Harbor Laboratory, the European Molecular Biology Organization (EMBO), and Rockefeller University. It was announced in December 2017 and will be launched in March 2018. The editors of *LSA* will consider directly submitted papers, but also the best of the 10,000 submissions that were declined by nine highly selective journals published by the three partner organizations. This "cascade" publishing model has become important to research

Announcement of *Life Science Alliance (LSA)* with representatives from Cold Spring Harbor Laboratory, EMBO, and Rockefeller University, at the American Society for Cell Biology meeting, December 2017.

communication in recent years, benefiting authors whose papers find an appropriate venue for publication more efficiently. *LSA* is a unique cross-publisher cascade and may attract other partners.

In the book-publishing program, 15 new print titles and 18 new e-books were added in 2017. *Malaria: Biology in the Era of Eradication* was generously supported by the J.C. Flowers Foundation in the name of the Isdell:Flowers Cross Border Malaria Initiative. This support enabled us to provide the book free of charge to scientists, clinicians, and public health officials in low-resource, malaria-endemic countries throughout the world in a format readable on every smart phone, tablet, e-reader, or computer. The book is downloadable from our website and licensed for open redistribution. In addition, thanks to the generosity of the book's editors (who waived their honoraria), free print copies were made available to academic libraries in malaria-endemic areas and to individual scientists and health workers in the United States and elsewhere. The third edition of the now-classic *Essentials of Glycobiology* was published in August, along with a free online edition available from the National Center for Biotechnology Information's Bookshelf. The year's best-selling new book was *A Cure Within: Scientists Unleashing the Immune System to Kill Cancer*, by Neil Canavan, an absorbing series of conversations with the pioneers of immunotherapy intended for the nonspecialist reader.

Japanese translation of *Navigating Metabolism* and Chinese translation of *Molecular Cloning,* Fourth Edition

A Chinese translation of *Molecular Cloning*, Fourth Edition, and a Japanese translation of *Navigating Metabolism* were recently completed. A Japanese translation of *Orphan: The Quest to Save Children with Rare Genetic Disorders* and a Chinese translation of *A Cure Within* are currently under way.

The strong performance of our book program in 2017 was assisted by the continued expansion of direct-to-customer e-book offerings. The popularity of e-books from third-party sources such as Amazon's Kindle store has been low, and the new program's objective is to add value to print books

CSHL Press exhibit booth at ASCB, December 2017.

bought directly from the Press and thus compete more effectively with commercial suppliers. The 2017 e-book revenue helped maintain direct sales from the Press website at >20% of total book sales, an exceptional result in today's online-dominated retail environment. Sixty-one percent of purchases included an e-book, and many generated requests for direct notification of publication news and special offers from the Press.

Staff

During the year, we welcomed Dana Macciola, Administrative Assistant, *Genome Research*, *Learning & Memory*, and *Molecular Case Studies*; Miriam Fein, Reviews Editor, *Genes & Development*; and Jennifer DeLeon, Assistant Editor, *Genome Research.*

We also said farewell with gratitude to Laura DeMare, Assistant Editor, *Genome Research* and Associate Editor, *Molecular Case Studies*.

The mission of the Press is to create publications and services that help scientists succeed while supporting the Laboratory's reputation and operating revenue. This requires engagement with many of the world's most accomplished scientists, and these relationships are sustained by the professional skills and dedication of the Press staff. I thank them all and recognize in particular those individuals who provide outstanding leadership in our diverse activities: Assistant Director Richard Sever, journal editors Terri Grodzicker and Hillary Sussman, and departmental directors Jan Argentine, Wayne Manos, Stephen Nussbaum, Marcie Siconolfi, and Linda Sussman. And, as always, I am extremely grateful for the patience, warmth, and efficiency of my Executive Assistant, Mala Mazzullo.

John Inglis
Executive Director and Publisher

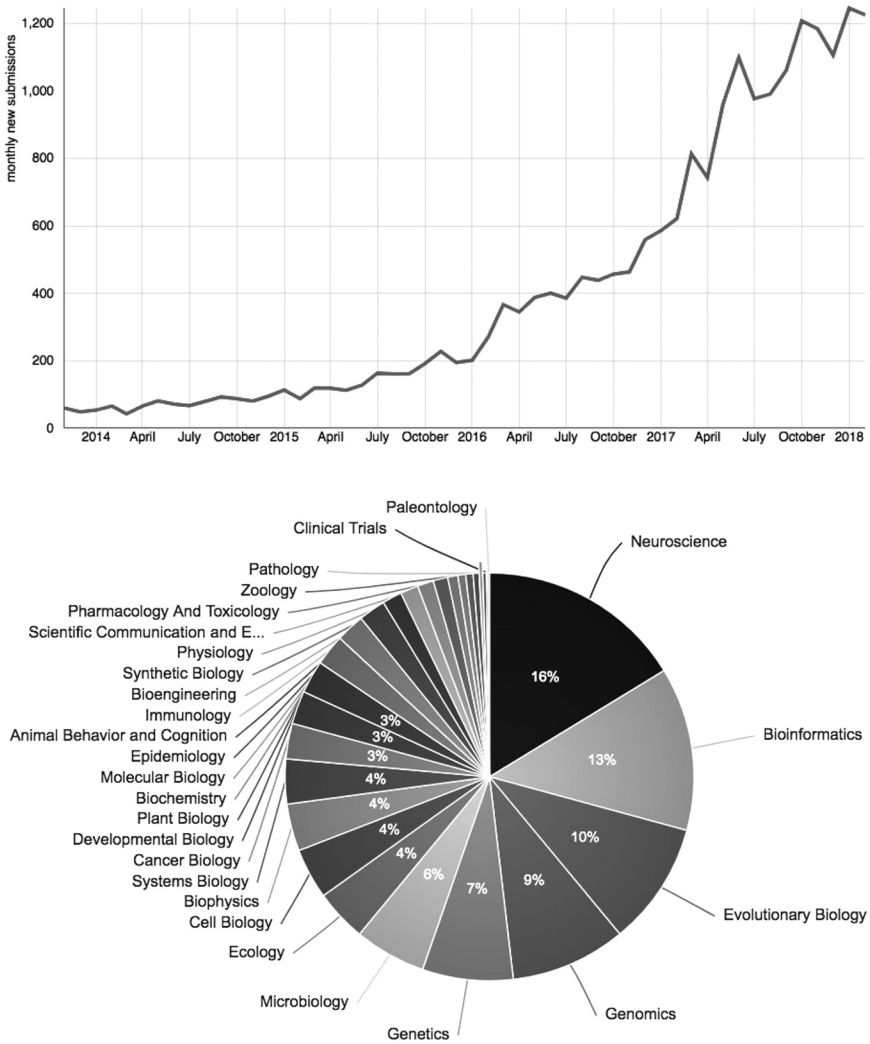
PREPRINT SERVER

bioRxiv

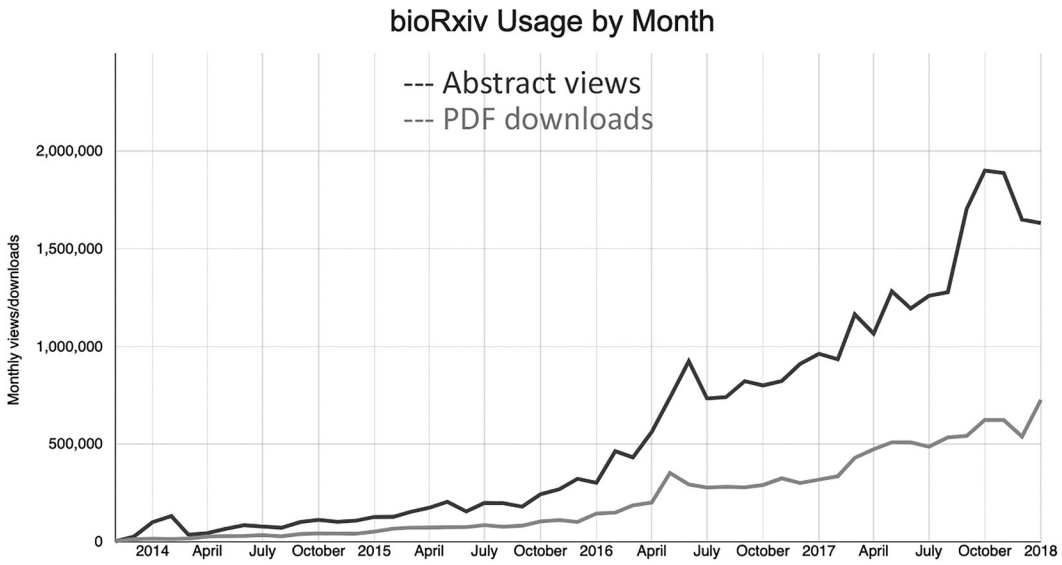
Preprints are research manuscripts yet to be certified by peer review and accepted by a journal. A preprint server is an online platform dedicated to the distribution of preprints.

bioRxiv is the not-for-profit preprint server for the life sciences that has been operated by Cold Spring Harbor Laboratory since November 2013. After start-up support from the Laboratory and vital additional funding from CSHL Trustee Robert Lourie, bioRxiv was fortunate to receive substantial, multiyear support from Chan Zuckerberg Initiative (CZI) in May 2017.

In 2017, the number of manuscripts on bioRxiv increased substantially and continually. More than 11,000 new submissions were received—more than double the previous year's total—to



(*Top*) Graph showing growth in new submissions. (*Bottom*) Pie chart showing percentage of bioRxiv submissions by category.



Site usage is growing.

bring the total content to nearly 19,000 manuscripts. Twenty-nine percent of the manuscripts on the server were revised at least once. Submissions originated from 7000 institutions in 100 countries. The most prolific are Stanford University, the universities of Cambridge, Oxford, and Washington, and Harvard University. The largest subject category was neuroscience.

Usage of the site, as seen in page views and article downloads, also grew impressively in 2017.

Nine percent of the manuscripts on bioRxiv have public comments, ranging from minor observations to substantial critique. Authors reported receiving more feedback privately, through e-mail or in person, than publicly.

bioRxiv's growing importance to science has prompted interest from individuals and organizations in the creation of platforms for commentary on preprints. Several of the earliest (PreLights, biOverlay, PREreview) point directly to bioRxiv content.

An author can transfer a bioRxiv manuscript directly, simply, and quickly to a journal for editorial consideration. In 2017, an author's choices for transfer rose to 120 journals from 31 publishers, resulting in more than 1000 transfers. Certain journals also give an author the opportunity to deposit a manuscript on bioRxiv immediately after submission to the journal, a convenience for authors that also increases the volume of content on bioRxiv.

Many bioRxiv preprints received attention in news outlets and were extensively discussed in social media. The development of bioRxiv itself was featured in many publications, including *Nature* and *Science*. In a recent study, 18% of bioRxiv preprints were found to have attention scores that put them in the top 5% of all papers appearing in the study period. Articles with a preprint have been shown to have higher attention scores and more citations when published than those without. Citations to bioRxiv manuscripts themselves in the peer-reviewed literature are increasingly frequent, but standardization in preprint citation practice is lacking.

bioRxiv's growth and community enthusiasm for preprints have caught the attention of scientists, funding agencies, and publishers, and there is widespread speculation about the consequences for journals and the assessment of research. bioRxiv's founders receive many invitations to speak at academic institutions and participate in conference presentations and panel discussions. bioRxiv's progress has prompted preprint initiatives in many other disciplines, including chemistry, earth and space sciences, psychology, and sociology.

Staff

The year 2017 saw a widening recognition that bioRxiv is transforming how life scientists communicate. The generous support of CZI made hiring full-time staff possible. Kevin-John (KJ) Black, a software developer with Silicon Valley experience and deep knowledge of bioRxiv's submission system and hosting platform, joined as Product Lead in September. Dr. Samantha Hindle, an open science advocate, was appointed in December as Content Lead after postdoctoral research at UCSF. Sam and KJ are ideal team members and have greatly increased bioRxiv's capacity. Four experienced, part-time freelance editors assist in the first phase of manuscript screening. The second phase relies on a growing cadre of affiliate scientists—principal investigators who assist in screening manuscripts and provide a sounding board on many issues relevant to bioRxiv's development. The bioRxiv team also benefits enormously from the help and advice of Cold Spring Harbor Laboratory Press staff. The collective efforts of all involved ensure that most bioRxiv manuscripts, remarkably, post less than a day after submission.

John Inglis
Co-Founder, bioRxiv

FINANCE

FINANCIAL STATEMENTS

CONSOLIDATED BALANCE SHEET
December 31, 2017
(with comparative financial information as of December 31, 2016)

		2017	2016
Assets:			
Cash and cash equivalents	$	64,179,016	58,209,476
Grants receivable		8,948,017	8,066,970
Contributions receivable, net		74,760,266	47,353,618
Investments		531,826,534	472,977,094
Investment in employee residences		6,348,606	6,000,849
Restricted use assets		4,070,570	3,522,055
Other assets		11,190,405	6,323,460
Land, buildings, and equipment, net		235,245,898	229,093,660
Total assets	$	936,569,312	831,547,182
Liabilities and net assets:			
Liabilities:			
Accounts payable and accrued expenses	$	15,888,844	10,030,041
Deferred revenue		5,601,953	9,125,481
Interest rate swap		31,345,495	32,713,773
Bonds payable		95,741,427	95,675,157
Total liabilities		148,577,719	147,544,452
Net assets:			
Unrestricted		403,766,549	364,233,745
Temporarily restricted		266,162,378	205,772,108
Permanently restricted		118,062,666	113,996,877
Total net assets		787,991,593	684,002,730
Total liabilities and net assets	$	936,569,312	831,547,182

CONSOLIDATED STATEMENT OF ACTIVITIES
Year ended December 31, 2017
(with summarized financial information for the year ended December 31, 2016)

	Unrestricted	*Temporarily Restricted*	*Permanently Restricted*	*2017 Total*	*2016 Total*
Revenue and other support:					
Public support—contributions and nonfederal grant awards	$ 24,748,370	56,086,908	4,065,789	84,901,067	69,813,418
Federal grant awards	34,668,573	–	–	34,668,573	32,371,500
Indirect cost allowances	31,719,355	–	–	31,719,355	29,781,859
Investment return utilized	21,297,672	–	–	21,297,672	18,455,646
Royalty and license revenue	13,399,961	–	–	13,399,961	3,290,181
Program fees	8,741,475	–	–	8,741,475	7,444,287
Publications sales	9,411,102	–	–	9,411,102	9,567,069
Dining services	4,993,409	–	–	4,993,409	4,596,996
Rooms and apartments	3,922,034	–	–	3,922,034	3,801,520
Miscellaneous	663,437	–	–	663,437	687,760
Net assets released from restrictions	18,202,531	(18,202,531)	–	–	–
Total revenue and other support	171,767,919	37,884,377	4,065,789	213,718,085	179,810,236
Expenses:					
Research	96,658,762	–	–	96,658,762	90,381,188
Educational programs	18,392,576	–	–	18,392,576	18,449,272
Publications	9,223,370	–	–	9,223,370	8,880,801
Banbury Center conferences	1,598,803	–	–	1,598,803	1,545,507
DNA Learning Center programs	2,169,075	–	–	2,169,075	2,385,791
Watson School of Biological Sciences programs	3,111,700	–	–	3,111,700	2,949,461
General and administrative	18,362,195	–	–	18,362,195	17,879,756
Dining services	6,471,584	–	–	6,471,584	5,949,534
Total expenses	155,988,065	–	–	155,988,065	148,421,310
Excess of revenue and other support over expenses	15,779,854	37,884,377	4,065,789	57,730,020	31,388,926
Other changes in net assets:					
Investment return excluding amount utilized	22,384,672	22,505,893	–	44,890,565	2,738,316
Change in fair value of interest rate swap	1,368,278	–	–	1,368,278	1,338,359
Increase in net assets	39,532,804	60,390,270	4,065,789	103,988,863	35,465,601
Net assets at beginning of year	364,233,745	205,772,108	113,996,877	684,002,730	648,537,129
Net assets at end of year	$ 403,766,549	266,162,378	118,062,666	787,991,593	684,002,730

CONSOLIDATED STATEMENT OF CASH FLOWS
Year ended December 31, 2017
(with comparative financial information for the year ended December 31, 2016)

	2017	2016
Cash flows from operating activities:		
Increase in net assets	$ 103,988,863	35,465,601
Adjustments to reconcile increase in net assets to net cash provided by operating activities:		
Change in fair value of interest rate swap	(1,368,278)	(1,338,359)
Depreciation and amortization	13,441,446	13,694,691
Amortization of deferred bond costs	66,270	66,270
Net appreciation in fair value of investments	(62,254,634)	(18,048,396)
Contributions restricted for long-term investment	(12,059,874)	(10,453,798)
Changes in assets and liabilities:		
Grants receivable	(881,047)	1,478,385
Contributions receivable, net	(23,650,821)	(10,729,807)
Restricted use assets	(548,515)	1,890,048
Other assets	(4,866,945)	3,947,423
Accounts payable and accrued expenses, net of financing activities	3,746,487	481,895
Deferred revenue	(3,523,528)	1,120,839
Net cash provided by operating activities	12,089,424	17,574,792
Cash flows from investing activities:		
Capital expenditures	(19,593,684)	(12,168,371)
Proceeds from sales and maturities of investments	100,074,927	40,841,045
Purchases of investments	(96,669,733)	(45,837,750)
Net change in investment in employee residences	(347,757)	160,554
Net cash used in investing activities	(16,536,247)	(17,004,522)
Cash flows from financing activities:		
Contributions restricted for long-term investment	4,065,789	883,150
Contributions restricted for investment in capital	7,994,085	9,570,648
Increase in contributions receivable	(3,755,827)	(6,523,768)
Increase (decrease) in accounts payable relating to capital expenditures	2,112,316	(500,000)
Net cash provided by financing activities	10,416,363	3,430,030
Net increase in cash and cash equivalents	5,969,540	4,000,300
Cash and cash equivalents at beginning of year	58,209,476	54,209,176
Cash and cash equivalents at end of year	$ 64,179,016	58,209,476
Supplemental disclosure:		
Interest paid	$ 3,886,138	3,816,392

FINANCIAL SUPPORT OF THE LABORATORY

Cold Spring Harbor Laboratory, Banbury Center, and the Dolan DNA Learning Center receive a substantial portion of funding through grants from the federal government, and through grants, capital gifts, and annual contributions from New York State, private foundations, corporations, and individuals. The following section summarizes funding that occurred during 2017.

GRANTS January 1–December 31, 2017

COLD SPRING HARBOR LABORATORY

Grantor	*Program/Principal Investigator*	*Duration of Grant*		*2017 Funding*[1]	
FEDERAL GRANTS NATIONAL INSTITUTES OF HEALTH					
Program Project and Center Support	Drs. Stillman/Krainer/D. Spector/Vakoc	05/25/12	12/31/17	2,165,000	
	Dr. Tuveson—Cancer Center Core	08/01/16	07/31/21	4,395,404	
	Drs. McCombie/Tuveson	09/25/15	08/31/19	154,908	
Cooperative Research Agreement Support[2]	Dr. Gingeras	09/21/12	07/31/18	689,972	
	Drs. Huang/Gillis/Mitra/Osten/Zador	09/21/17	06/30/22	4,846,662	*
	Dr. Kepecs	09/30/15	08/31/18	390,147	
	Drs. Krasnitz/Wigler	03/01/15	02/28/18	314,160	
	Drs. Osten/Albeanu/Mitra	09/20/17	06/30/22	1,880,281	*
	Dr. Tuveson	06/12/14	02/28/19	189,889	
	Dr. Zador	09/25/15	06/30/18	550,303	
Contract Support Leidos Biomedical Research, Inc.—NCI[2]	Drs. Tuveson/D. Spector	01/09/17	12/31/18	1,943,269	*
Research Support	Dr. Albeanu	07/01/13	06/30/18	480,000	
	Dr. Albeanu	02/09/16	01/31/21	408,000	
	Dr. Churchland	03/01/13	02/28/18	472,500	
	Drs. Dobin/Gingeras	08/18/17	05/31/22	480,000	*
	Dr. Furukawa	03/01/15	02/29/20	498,864	
	Dr. Furukawa	05/08/14	03/31/18	550,980	
	Dr. Gillis	09/15/17	08/31/21	464,956	*
	Drs. Gillis/Huang/Lee	07/13/17	05/31/22	472,477	*
	Dr. C. Hammell	03/01/16	12/31/20	403,200	
	Dr. Huang	07/01/13	03/31/18	741,089	
	Drs. Huang/Gillis	08/01/16	07/31/21	811,483	
	Drs. Huang/Wigler	03/01/14	01/31/19	721,116	
	Dr. Joshua-Tor	06/10/16	03/31/20	345,600	
	Dr. Kepecs	09/15/17	07/31/22	399,875	*
	Dr. Kepecs	04/01/14	03/31/19	480,000	
	Dr. Kepecs	08/15/15	05/31/20	420,000	
	Dr. Kepecs	09/15/15	07/31/20	438,772	
	Drs. Koulakov/Zador	07/01/13	03/31/18	432,000	
	Dr. Krainer	07/01/17	06/30/22	806,400	*

[1]Awarded, including direct and indirect costs
[2]Funding amounts include only CSHL's portion of the award
*New or competing renewals or supplements awarded in 2017

Grantor	*Program/Principal Investigator*	*Duration of Grant*		*2017 Funding*[1]	
	Dr. Lee	08/19/16	05/31/21	480,000	
	Dr. Li	03/01/14	02/28/19	480,000	
	Dr. Li	07/11/17	06/30/19	288,000	*
	Drs. Li/Huang	09/28/15	06/30/20	660,707	
	Dr. Martienssen	06/01/17	04/30/21	422,400	*
	Dr. Mills	09/16/14	08/31/19	634,545	
	Dr. Mitra	09/30/16	06/30/19	425,138	
	Dr. Park	09/15/16	08/31/21	174,117	
	Dr. Pedmale	08/04/17	07/31/22	480,000	*
	Dr. Shea	12/08/15	11/30/20	518,964	
	Dr. Sheltzer	09/18/15	08/31/20	480,000	
	Dr. Siepel	09/01/14	12/31/17	329,759	
	Dr. Siepel	04/24/15	01/31/19	192,000	
	Dr. D. Spector	04/01/15	03/31/19	766,080	
	Dr. Stenlund	05/01/15	04/30/20	504,989	
	Dr. Stillman	03/01/17	02/28/21	729,600	*
	Dr. Tonks	05/14/15	04/30/20	776,676	
	Dr. Tonks	01/15/15	12/31/18	448,737	
	Dr. Trotman	07/30/14	06/30/19	398,400	
	Drs. Tuveson/M. Hammell/Pappin	12/07/16	11/30/21	554,114	
	Drs. Tuveson/M. Hammell/Pappin/Trotman	09/22/14	08/31/18	552,684	
	Dr. Vakoc	04/01/13	03/31/19	557,762	
	Dr. Van Aelst	08/01/13	05/31/18	532,426	
	Dr. Van Aelst	04/01/13	03/31/19	489,466	
	Dr. Zador	07/01/14	04/30/19	420,000	
	Dr. Zador	09/01/17	08/31/20	907,874	*
Research Subcontracts					
NIH/Boston Children's Hospital Consortium Agreement	Dr. Mitra	09/23/15	06/30/18	60,316	
NIH/Emory University Consortium Agreement	Dr. Huang	04/01/14	02/28/18	37,800	
NIH/Harvard Medical School Consortium Agreement	Dr. Osten	07/01/17	04/30/22	84,000	*
NIH/Johns Hopkins University Consortium Agreement	Dr. Joshua-Tor	03/01/15	02/29/20	33,264	
NIH/New York Genome Center Consortium Agreement	Drs. Wigler/Iossifov/Levy/Siepel	01/14/16	11/30/17	233,565	
NIH/New York University Consortium Agreement	Dr. Koulakov	06/01/14	05/31/19	173,200	
NIH/The Research Foundation for the State of New York—Stony Brook Consortium Agreement	Dr. Wigler	05/01/14	04/30/19	130,599	
NIH/The Research Foundation for the State of New York—Stony Brook Consortium Agreement	Dr. M. Hammell	09/15/17	07/30/22	86,809	*
NIH/ The Scripps Research Institute Consortium Agreement	Dr. Tuveson	08/01/16	07/31/19	95,329	
NIH/University of Minnesota Consortium Agreement	Dr. dos Santos	01/01/17	12/31/20	100,345	*
NIH/University of Nebraska Consortium Agreement	Drs. Tuveson/Pappin	05/01/17	04/30/22	365,750	*
Fellowship / Career Development Support	Dr. Baker	02/01/15	10/31/17	47,190	
	Dr. Bravo-Rivera	06/01/17	05/31/20	56,694	*

[1]Awarded, including direct and indirect costs

*New or competing renewals or supplements awarded in 2017

Grantor	*Program/Principal Investigator*	*Duration of Grant*		*2017 Funding*[1]
	Dr. Crow	06/16/17	06/15/20	59,166 *
	Dr. Diermeier	04/07/17	03/31/19	112,407 *
	Dr. Engle	08/01/16	07/31/18	128,235
	Dr. Hwang	03/01/15	02/28/18	63,078
	Dr. Regan	12/01/15	11/30/18	64,114
	Dr. Wang	05/02/16	05/01/19	66,450
	Dr. Wong	08/01/15	07/31/18	59,166
Institutional Training Program Support	Dr. Mills/Cancer Postdoctoral	09/16/16	08/31/21	125,761
	Dr. Gann/Watson School of Biological Sciences	07/01/17	06/30/22	279,669 *
Course Support	Advanced Sequencing Technologies and Applications	04/10/12	06/30/18	70,160
	Advanced Techniques in Molecular Neuroscience	07/06/15	03/31/20	105,668
	Cell and Development Biology of *Xenopus*	05/05/14	03/31/19	76,284
	Cellular Biology of Addiction	08/01/16	07/31/21	51,299
	Chromatin, Epigenetics and Gene Expression	04/13/17	03/31/22	96,949 *
	Computational and Comparative Genomics	08/15/17	06/30/20	62,304 *
	Empowering Nextgen Advanced Biomedical Leadership	06/01/15	02/29/20	521,039
	Expression, Purification & Analysis and Protein Complexes	04/13/17	03/31/22	101,003 *
	Mouse Development, Stem Cells and Cancer	04/13/17	03/31/22	121,703 *
	Programming for Biology	09/01/17	06/30/20	83,641 *
	Quantitative Imaging: From Cells to Molecules	04/01/16	03/31/21	102,233
	Vision: A Platform for Linking Circuits, Perception and Behavior	04/01/15	03/31/18	30,000
	X-Ray Methods in Structural Biology	09/01/17	08/31/22	90,745 *
Meeting Support	The Biology of Genomes	04/01/13	03/31/18	54,600
	Cell Death	04/01/17	03/31/18	5,000 *
	Eukaryotic DNA Replication & Genome Maintenance	08/10/17	07/31/18	3,000 *
	Eukaryotic mRNA Processing Conference	07/01/17	06/30/18	4,000 *
	Genome Informatics	08/01/17	07/31/18	32,293 *
	Glia in Health and Disease	09/15/17	08/31/18	25,000 *
	Global Regulation of Gene Expression	02/22/16	01/31/19	29,951
	Mechanisms of Metabolic Signaling	04/17/17	03/31/18	15,000 *
	Microbial Pathogenesis and Host Response	08/11/17	07/31/18	6,500 *
	Neurobiology of *Drosophila*	09/01/17	08/31/18	20,000 *
	Neuronal Circuits	09/25/17	09/29/18	18,000 *
	Protein Homeostasis in Health and Disease	09/15/17	08/31/18	30,724 *
	Retroviruses	01/01/17	12/31/17	35,000 *
	Systems Biology: Networks	03/17/17	02/28/18	29,334 *
	Telomeres and Telomerase	04/01/17	03/31/18	18,000 *
	Wiring the Brain	04/01/17	03/31/18	10,000 *
NATIONAL SCIENCE FOUNDATION				
Multiple Project Award Support	Drs. Gingeras/Jackson/Martienssen/McCombie/Ware	06/15/16	05/31/19	2,145,722
	Drs. Jackson/Lippman	09/01/16	08/31/20	1,143,923

[1]Awarded, including direct and indirect costs

*New or competing renewals or supplements awarded in 2017

Grantor	*Program/Principal Investigator*	*Duration of Grant*		*2017 Funding*[1]	
Research Support	Dr. Albeanu	08/01/17	07/31/21	250,000	*
	Dr. Jackson	06/01/15	05/31/18	203,701	
	Dr. Kepecs	09/01/17	08/31/19	397,248	*
	Dr. Lippman	06/15/16	05/31/19	208,271	
	Dr. Lippman	07/01/17	06/30/21	1,269,271	*
	Dr. Lippman	11/01/14	10/31/17	215,164	
	Drs. McCombie/Martienssen	06/01/15	05/31/18	390,621	
	Dr. Siepel	03/01/16	02/28/19	90,695	
Research Subcontracts					
NSF—Columbia University Consortium Agreement	Dr. Kepecs	11/15/16	10/31/18	71,395	
NSF—Cornell University Consortium Agreement	Dr. Timmermans	02/01/13	03/31/18	277,830	
NSF—Cornell University Consortium Agreement	Dr. Ware	05/15/13	04/30/18	160,786	
NSF—Iowa State University Consortium Agreement	Dr. Jackson	03/01/13	02/28/18	399,788	
NSF—University of Arizona Consortium Agreement	Dr. Ware/Micklos	09/01/13	08/31/18	1,116,738	
Fellowship Support	B. Bibel	09/01/14	07/31/19	46,000	
	B. Berube	06/01/17	05/31/20	46,000	*
	D. Johnson	06/01/17	05/31/20	46,000	*
	Dr. Lemmon (Direct)	07/01/15	06/30/18	72,000	
Institutional Training Program Support	Drs. Churchland/C. Hammell	05/01/16	04/30/19	153,193	
Course Support	Advanced Bacterial Genetics	06/15/17	05/31/22	90,000	*
	Drosophila Neurobiology: Genes, Circuits, and Behavior	07/01/17	06/30/20	25,000	*
	Frontiers and Techniques in Plant Science	05/01/15	04/30/18	107,314	
	Synthetic Biology	07/01/15	06/30/18	33,300	
	Yeast Genetics and Genomics	07/01/17	06/30/22	90,000	*
UNITED STATES DEPARTMENT OF AGRICULTURE					
Research Support	Dr. Jackson	12/15/15	12/14/19	124,686	
	Dr. Lippman	01/01/15	12/31/18	124,018	
	Dr. Lippman	11/01/15	10/31/19	104,517	
	Dr. McCombie	09/15/17	09/14/18	868,303	*
UNITED STATES DEPARTMENT OF THE ARMY					
Research Support	Dr. Egeblad	06/15/14	06/14/19	945,000	
Research Subcontracts					
DOA/Emory University Consortium Agreement	Dr. Trotman	09/30/16	09/29/19	102,067	
DOA/University of Southern California Consortium Agreement	Dr. Churchland	08/23/16	11/22/18	94,500	
Fellowship Support	Dr. Casanova Salas	08/15/17	08/14/19	120,000	*

[1]Awarded, including direct and indirect costs

*New or competing renewals or supplements awarded in 2017

Grantor	*Program/Principal Investigator*	*Duration of Grant*		*2017 Funding*[1]
UNITED STATES DEPARTMENT OF ENERGY				
Research Support	Dr. Martienssen	09/15/17	09/14/21	1,151,836 *
Research Subcontracts				
DOE/Lawrence Berkeley National Laboratory Consortium Agreement	Dr. Ware	05/18/17	09/30/20	649,793 *
DOE/New York University Consortium Agreement	Dr. McCombie	08/15/15	08/14/20	236,022
UNITED STATES DEPARTMENT OF THE INTERIOR				
Research Subcontracts				
DOI/Harvard University Consortium Agreement	Dr. Zador	01/15/16	01/14/21	1,865,261
MISCELLANEOUS SOURCES OF FUNDING				
Equipment Support				
CSHL Translational Cancer Support	Dr. Osten	07/21/17	07/20/20	277,147 *
	Dr. Preall	06/01/17	05/31/20	398,167 *
Howard Hughes Medical Institute	Dr. Joshua-Tor	12/04/17	12/03/18	350,000 *
Program Project Support				
The Simons Foundation/Autism	Dr. Wigler	01/01/17	12/31/20	17,57,782 *
The Simons Foundation/Cancer	Dr. Wigler	01/01/17	12/31/19	2,526,074 *
Research Support				
Dr. Dinu F. Albeanu	Dr. Albeanu	06/27/17	06/26/18	8,000 *
Rita Allen Foundation	Dr. dos Santos	09/01/16	08/31/21	100,000
	Dr. M. Hammell	09/01/14	08/31/19	110,000
	Dr. Zador	12/01/17	11/30/18	250,000 *
American Association for Cancer Research	Dr. Vakoc	07/01/16	06/30/18	100,000
American Cancer Society	Dr. Trotman	07/01/14	06/30/18	198,000
American Lung Association	Dr. Van Aelst	07/01/16	06/30/18	100,000
Anonymous	Cold Spring Harbor Laboratory	05/01/17	04/30/22	1,091,753 *
Beckman Research Institute of the City of Hope	Dr. Atwal	06/01/15	06/30/18	200,000
Boehringer Ingelheim RCV GmbH & Co KG	Dr. Vakoc	09/15/15	09/14/18	862,695
Brain & Behavior Research Foundation	Dr. C. Hammell	09/15/17	09/14/19	50,000 *
Breast Cancer Alliance	Dr. Sheltzer	02/01/17	06/30/19	62,500 *
The Breast Cancer Research Foundation	Drs. Wigler/Levy/Mitra	10/01/17	09/30/18	250,000 *
Vincent & Patricia Breitenbach	Dr. Krainer	02/28/17	02/27/18	10,000 *
Cedar Hill Foundation	Dr. Fearon	11/16/17	11/15/18	60,000 *
Chen & Shao Anti-Cancer Foundation	Dr. Zheng	04/02/17	04/01/18	100,000 *
The Children's Heart Foundation	Dr. Ronemus	01/01/17	12/31/17	50,000 *
Collaborative Center for X-Linked Dystonia Parkinsonism of Massachusetts General Hospital	Drs. Lyon/Gillis	01/01/16	06/30/18	275,000

[1]Awarded, including direct and indirect costs

*New or competing renewals or supplements awarded in 2017

Grantor	*Program/Principal Investigator*	*Duration of Grant*		*2017 Funding*[1]	
Columbia University/The Simons Foundation	Dr. Krainer	09/01/14	08/31/17	58,618	
CSHL Translational Cancer Support	Dr. Atwal	02/25/17	02/24/20	335,419	*
	Dr. Atwal	07/17/17	07/16/18	111,760	*
	Dr. Chang	06/01/15	05/31/18	119,340	
	Dr. dos Santos	05/26/17	05/25/20	288,000	*
	Dr. Egeblad	06/15/16	06/14/19	429,488	
	Dr. Egeblad	03/01/17	02/28/18	70,495	*
	Dr. Gingeras	01/15/16	01/14/19	506,426	
	Dr. Kepecs	07/01/17	06/30/20	135,982	*
	Dr. Kinney	02/25/17	02/24/21	457,122	*
	Dr. Lee	06/01/15	05/31/18	350,620	
	Dr. Lyons	07/01/17	06/30/20	294,306	*
	Dr. Lyons	06/01/15	05/31/18	433,221	
	Dr. Mills	01/01/17	12/31/20	576,000	*
	Dr. Mills	10/25/17	10/24/21	701,131	*
	Dr. Mitra	07/01/17	06/30/20	348,763	*
	Dr. Osten	07/21/17	07/20/20	146,819	*
	Dr. Pappin	05/26/17	05/25/20	697,453	*
	Dr. Preall	06/01/17	05/31/20	977,914	*
	Dr. Sheltzer	05/26/17	05/25/18	337,751	*
	Dr. Sordella	04/27/17	04/26/19	501,857	*
	Dr. D. Spector	06/01/15	05/31/18	313,663	
	Dr. D. Spector	06/01/16	05/31/21	189,615	
	Dr. Tiriac	03/01/17	02/28/18	31,296	*
	Dr. Trotman	06/01/16	05/31/19	490,937	
	Dr. Tuveson	06/01/15	05/31/18	567,491	
	Dr. Vakoc	01/15/16	01/14/19	1,255,578	
	Dr. Van Aelst	06/01/15	05/31/20	378,758	
	Dr. Yeh	06/15/16	06/14/19	763,066	
	Dr. Yeh	07/01/17	06/30/20	28,007	*
	Dr. Zhang	01/13/17	01/12/21	349,476	*
	Dr. Zheng	06/01/15	05/31/19	395,222	
Donaldson Charitable Trust	Dr. Tuveson	12/21/16	12/20/19	267,244	
East–West International BV	Dr. Lippman	06/01/16	05/31/19	100,000	
William Guy Forbeck Research Foundation	Dr. Vakoc	07/01/16	06/30/18	80,000	
Glen Cove C.A.R.E.S.	Dr. Egeblad	01/28/17	01/27/18	7,000	*
The GoGo Foundation	Dr. Li	09/29/16	09/28/19	40,000	
Gyeongsang National University/ The Republic of Korea	Dr. Jackson	01/04/16	12/31/17	52,393	
Irving Hansen Foundation	Dr. Tonks	07/21/17	07/20/18	30,000	*
Jo-Ellen and Ira Hazan	Dr. Tuveson	12/18/17	12/17/18	2,500	*
Robert and Cindy Higginson	Dr. Mills	12/17/17	12/16/18	2,000	*
Human Frontier Science Program	Dr. Li	09/01/16	08/31/19	100,000	
Indian Institute of Technology Madras	Dr. Mitra	01/01/15	12/31/19	27,327	
The Kavli Foundation	Dr. Churchland	11/15/17	10/31/18	35,000	*
Eugenia and Thomas Korossy	Dr. Zhang	07/20/17	07/19/18	15,000	*
Marilyn and Michael Leidner	Dr. D. Spector	04/29/16	04/28/18	450	
The Leukemia & Lymphoma Society	Dr. Vakoc	07/01/15	06/30/20	110,000	
Long Island Bioscience Hub	Dr. Vaughan	11/15/17	11/14/18	100,000	*
The Lustgarten Foundation	Dr. Fearon	07/01/14	12/31/19	1,000,000	
	Dr. Tuveson	09/01/17	08/31/22	1,000,000	*
	Dr. Tuveson	06/15/12	06/30/18	70,000	
	Drs. Tuveson/Krasnitz	05/01/16	04/30/18	1,348,645	

[1]Awarded, including direct and indirect costs

*New or competing renewals or supplements awarded in 2017

Grantor	*Program/Principal Investigator*	*Duration of Grant*		*2017 Funding*[1]
The G. Harold and Leila Y. Mathers Charitable Foundation	Dr. Mitra	01/01/17	12/31/19	330,000 *
Breast Cancer Awareness Day in Memory of Elizabeth McFarland	Dr. Wigler	01/01/16	12/31/18	43,658
The Don Monti Memorial Research Foundation	Drs. Stillman/Tonks/Vakoc	07/20/17	07/19/18	150,000 *
Louis Morin Charitable Trust	Dr. Gingeras	12/12/17	12/11/18	125,000 *
Memorial Sloan Kettering Cancer Center /Henry and Marilyn Taub Foundation	Dr. Krainer	12/15/16	12/14/19	133,333
Natures Answer Foundation	Dr. dos Santos	12/14/17	12/13/18	5,000 *
New York State Department of Health	Dr. Egeblad	01/01/17	12/31/18	175,400 *
In Memory of Jack O'Brien	Dr. Hicks	08/01/14	12/31/17	3,000
Omicia, Inc.	Dr. Lyon	11/17/14	02/02/18	5,000
The Michelle Paternoster Foundation	Dr. Vakoc	09/16/15	09/15/18	50,000
The Pershing Square Foundation	Dr. Egeblad	07/01/17	06/30/20	200,000 *
	Dr. Vakoc	09/09/16	07/31/19	200,000
The Pew Charitable Trusts	Dr. Churchland	08/01/14	07/31/19	75,000
Christina Renna Foundation Inc.	Dr. Vakoc	09/16/15	09/15/18	30,000
Ride for Life, Inc.	Dr. M. Hammell	12/20/17	12/19/18	10,000 *
RIKEN	Dr. Mitra	04/01/15	03/31/18	2,809
Charles and Marie Robertson Foundation	Dr. dos Santos	01/09/17	01/08/18	40,000 *
Diane Emdin Sachs Memorial Fund	Dr. Sordella	09/01/12	08/31/17	135
Seven Bridges Genomics Inc.	Dr. Lyon	09/01/16	08/31/20	1,000
The Simons Foundation	Dr. Churchland	07/01/17	06/30/22	140,400 *
	Dr. Iossifov	05/01/17	04/30/19	177,830 *
	Dr. Zador	07/01/17	06/30/22	70,200 *
	Dr. Zador	01/01/17	12/31/18	150,000 *
The Simons Foundation/CSHL Innovative Cancer Research	Dr. Stillman	07/01/17	06/30/19	
	Drs. dos Santos/Egeblad/Van Aelst	07/01/17	06/30/19	659,980 *
	Dr. Lyons	07/01/17	06/30/20	397,797 *
	Drs. Lyons/Trotman	07/01/17	06/30/20	222,244 *
	Dr. Tonks	07/01/17	06/30/20	657,191 *
	Drs. Tuveson/Vakoc	07/01/17	06/30/20	250,000 *
Starr Cancer Consortium	Dr. Joshua-Tor	01/01/17	12/31/18	142,200 *
	Drs. Lee/M. Hammell	01/01/16	06/30/18	298,800
	Dr. Vakoc	01/01/16	12/31/18	149,400
	Dr. Wigler	01/01/16	12/31/17	49,800 *
Theresa Steen	Dr. Atwal	04/01/17	03/31/18	100 *
Three Strohm Sisters Family Foundation	Dr. Egeblad	12/13/17	12/12/18	5,000 *
The Swartz Foundation	Dr. Zador	01/01/17	12/31/17	53,000 *
Swim Across America	Dr. Sordella	12/26/17	12/25/18	60,000 *
Tackle Autism	Dr. Wigler	12/14/17	12/13/18	5,000 *
Anne D. Thomas	Dr. dos Santos	08/08/17	08/07/18	10,000 *
The Thompson Family Foundation, Inc.	Drs. Tuveson/Chang/Egeblad/Fearon/Lyons/ Vakoc/Yeh	03/06/17	03/05/21	1,365,410 *
Friends of the TJ Foundation Inc.	Dr. Vakoc	09/16/15	09/15/18	50,000
U.S.–Israel Binational Agricultural Research and Development Fund	Dr. Lippman	12/01/15	11/30/18	47,000
	Dr. Shea	09/01/16	08/31/20	16,667

[1]Awarded, including direct and indirect costs

*New or competing renewals or supplements awarded in 2017

Grantor	*Program/Principal Investigator*	*Duration of Grant*		*2017 Funding*[1]	
University of Iowa/The Simons Foundation	Dr. Mills	01/01/17	08/31/18	45,373	*
University of Southern California/ The Susan G. Komen Breast Cancer Foundation	Dr. Levy	03/21/17	10/21/18	27,152	*
The V Foundation	Dr. Atwal	12/19/16	12/18/18	83,719	
	Dr. dos Santos	11/01/16	11/01/18	100,000	
	Dr. Tuveson	11/01/16	11/01/19	200,000	
The Wasily Family Foundation	Dr. Tuveson	06/27/17	06/26/18	50,000	*
Joan & Sanford I. Weill Medical College	Dr. Fearon	07/01/14	06/30/19	78,520	
Women's Partnership in Science	Dr. Sordella	01/01/17	12/31/18	79,115	*
	Dr. Timmermans	01/01/17	12/31/18	27,999	*
	Dr. Van Aelst	01/01/17	12/31/18	80,000	*
The Bradley Zankel Foundation, Inc.	Dr. Zheng	01/21/17	01/20/18	15,000	*
Fellowship Support					
American–Italian Cancer Foundation	Dr. Tonelli	08/01/16	07/31/18	40,000	
Rita Allen Foundation	Dr. Fearon	01/01/17	12/31/17	6,000	*
Arnold and Mabel Beckman Foundation	Dr. Mejia	09/01/17	08/31/18	68,467	*
Brain and Behavior Foundation	Dr. Pi	01/15/17	01/14/19	35,000	*
	Dr. Wang	01/15/17	01/14/19	35,000	*
Terri Brodeur Breast Cancer Foundation	Dr. Albrengues	01/01/16	12/31/17	50,000	
European Molecular Biology Organization	Dr. Furlan	07/01/17	06/30/19	44,291	*
Genentech Foundation	Watson School of Biological Sciences	10/01/15	09/30/18	75,553	
Lola A. Goldring	Dr. Stillman	10/01/17	09/30/18	100,000	*
Human Frontier Science Program	Dr. Biffi	06/01/15	05/31/18	55,620	
	Dr. Carnevale	01/01/16	12/31/18	53,640	
	Dr. Xu	04/01/16	03/31/19	53,640	
Annette Kade Charitable Trust	Watson School of Biological Sciences	12/27/17	12/26/18	25,000	*
The Susan G. Komen Breast Cancer Foundation, Inc.	Dr. Albrengues	11/29/16	11/28/19	60,000	
	Dr. Arun	12/18/15	12/17/18	60,000	
	Dr. Diermeier	09/07/16	12/20/17	60,000	
Natural Sciences and Research Council of Canada	Dr. Bonham	07/01/15	06/30/17	16,746	
The Patrina Foundation	Women in Science and Engineering Initiative	09/27/17	09/26/18	10,000	*
The Pew Charitable Trusts	Dr. Rodriguez-Leal	08/01/16	07/31/20	30,000	
John and Amy Phelan Foundation	Watson School of Biological Sciences	09/01/17	08/31/22	100,000	*
The Research Foundation for State University of New York–Stony Brook	B. Alagesan	04/15/16	04/14/20	4,200	
	D. Cheng	01/16/17	01/15/21	4,200	*
	J. Levine	09/15/15	09/14/19	4,200	
	A. Yu	07/16/17	07/15/19	4,200	*
Charles and Marie Robertson Foundation	Drs. Stephenson-Jones/Bravo-Rivera	01/09/17	01/08/18	15,000	*
Jason Sheltzer and Joan Smith	Women in Science and Engineering Initiative	12/15/17	12/14/18	669	*
The Simons Foundation	Dr. Chen	03/01/15	02/28/18	81,150	
	Dr. Kaufman	03/01/15	02/28/18	81,150	

[1]Awarded, including direct and indirect costs

*New or competing renewals or supplements awarded in 2017

Grantor	*Program/Principal Investigator*	*Duration of Grant*		*2017 Funding*[1]
Lauri Strauss Leukemia Foundation, Inc.	Dr. Tarumoto	12/20/16	12/19/17	10,000
Manoj & Ranjeeta Subudhi	Dr. Sturgill	04/17/17	04/16/18	600 *
The Swartz Foundation	Drs. Albeanu/Gupta	01/01/17	12/31/17	60,000 *
	Drs. Albeanu/Bast	01/01/17	12/31/17	60,000 *
	Drs. Koulakov/Shuvaev	01/01/17	12/31/17	51,000 *
Swiss National Science Foundation	Dr. Musall	08/01/17	01/31/19	23,000 *
U.S.–Israel Binational Agricultural Research and Development Fund	Dr. Hendelman	08/01/16	07/31/18	22,500
University of Southern California	N. Anaparthy	05/01/17	12/31/17	21,889 *
Training Support				
Katya H. Davey	Undergraduate Research Program	08/16/17	08/15/18	25,000 *
Lita Annenberg Hazan Foundation	Watson School of Biological Sciences	05/01/08	04/30/18	10,000
New York State Department of Economic Development	Dr. Mills/CSHL Cancer Gene Discovery & Post Doctoral Research Training Program	02/01/16	01/31/21	140,000
William Townsend Porter Foundation	Undergraduate Research Program	04/01/17	03/31/18	11,800 *
University of Notre Dame	Undergraduate Research Program	04/01/16	03/31/21	20,000
Course Support				
American Brain Tumor Association	Brain Tumors	12/01/17	11/30/18	27,110 *
Arvis Inc.	Quantitative Imaging: From Acquisition to Analysis	03/30/17	03/29/18	1,500 *
The Leona M. & Harry B. Helmsley Charitable Trust	Course Program	02/01/15	01/31/18	700,000
Howard Hughes Medical Institute	Course Program	08/01/15	07/31/19	600,000
	Enhancing Diversity in Biomedical Research Through Immersive Training Experiences	11/01/17	10/31/18	42,000 *
The Lustgarten Foundation	Workshop on Pancreatic Cancer	04/26/17	04/25/18	20,000 *
Pancreatic Cancer Action Network, Inc.	Workshop on Pancreatic Cancer	04/26/17	04/25/18	10,000 *
Society for Neuroscience/ International Brain Research Organization	Summer Neuroscience Course	07/01/17	06/30/18	29,041 *
Meeting Support				
AbbVie Inc.	Fundamental Immunology and its Therapeutic Potential	04/14/17	04/13/18	2,500 *
Advanced Accelerator Applications	Annexins	05/08/17	05/07/18	3,000 *
Alnylam US, Inc.	Forty Years of RNA Splicing: From Discovery to Therapeutics	01/18/17	01/17/18	25,000 *
	RNA & Oligonucleotide Therapeutics	12/01/16	12/20/17	10,000
Amgen	Forty Years of RNA Splicing: From Discovery to Therapeutics	01/18/17	01/17/18	25,000 *
Aniara Diagnostica LLC	Annexins	05/08/17	05/07/18	5,000 *
Annexin Pharmaceuticals AB	Annexins	05/08/17	05/07/18	5,000 *
Arcturus Therapeutics Inc.	RNA & Oligonucleotide Therapeutics	12/01/16	12/20/17	3,000
AstraZeneca	Cell Death	03/01/17	02/28/18	2,500 *
Biogen MA, Inc.	Neurodegenerative Diseases: Biology and Therapeutics	12/20/17	12/19/18	5,000 *
Biogen US	Forty Years of RNA Splicing: From Discovery to Therapeutics	01/18/17	01/17/18	50,000 *
Bristol–Myers Squibb Company	Annexins	05/11/17	05/10/18	5,100 *
Burroughs Wellcome Fund	Microbial Pathogenesis and Host Response	05/15/17	05/14/18	10,000 *

[1]Awarded, including direct and indirect costs

*New or competing renewals or supplements awarded in 2017

Grantor	*Program/Principal Investigator*	*Duration of Grant*		*2017 Funding*[1]	
Cleveland Clinic	STATs: Importance in Basic and Clinical Cancer Research	08/24/17	08/23/18	10,000	*
CSHL Translational Cancer Research	Biology of Cancer: Microenvironment, Metastasis and Therapy	01/13/17	01/12/18	113,613	*
	Cancer Therapies and Clinical Outcomes	01/13/17	01/12/18	78,387	*
Diagnostica Stago, Inc.	Annexins	05/08/17	05/07/18	500	*
Genentech, Inc.	Fundamental Immunology and its Therapeutic Potential	02/23/17	02/22/18	2,500	*
	Annexins	05/08/17	05/07/18	2,000	*
Ionis Pharmaceuticals, Inc.	Forty Years of RNA Splicing: From Discovery to Therapeutics	01/18/17	01/17/18	7,500	*
H. Lundbeck A/S	Wiring the Brain	12/01/16	05/31/17	12,000	
Moderna Therapeutics, Inc.	RNA & Oligonucleotide Therapeutics	12/01/16	12/20/17	15,000	
Joan and Sanford I. Weill Medical College & Graduate School of Medical Sciences of Cornell University	Annexins	05/08/17	05/07/18	10,000	*
Laboratory Disposable Products Inc.	Annexins	05/08/17	05/07/18	100	*
New England Biolabs, Inc.	Forty Years of RNA Splicing: From Discovery to Therapeutics	01/18/17	01/17/18	25,000	*
RNA Therapeutics	RNA & Oligonucleotide Therapeutics	12/01/16	12/20/17	15,000	
Rockefeller University	STATs: Importance in Basic and Clinical Cancer Research	08/24/17	08/23/18	25,000	*
ViiV Healthcare Company	Retrovirus	03/22/17	03/21/18	20,000	*
Library Support					
Anonymous		12/21/16	12/20/18	25,000	
The Ellen Brenner Memorial Fund		12/15/17	12/14/18	2,000	*
Mr. and Mrs. Philip Goelet		12/21/16	12/20/18	10,000	*
The New York State Education Department		07/01/17	06/30/18	4,181	*
Niels Nielsen		12/21/16	12/20/18	15,000	
Laboratory Press Support					
Ajit and Nissi Varki	Development costs of *Essentials of Glycobiology 3rd edition*	01/10/17	01/09/18	16,000	*

[1]Awarded, including direct and indirect costs

*New or competing renewals or supplements awarded in 2017

DNA LEARNING CENTER GRANTS

Grantor	*Program*	*Duration of Grant*	*2017 Funding*†
FEDERAL GRANTS			
National Institutes of Health	*Barcode Long Island*	7/14–3/19	279,365
National Science Foundation	*TRPGR; Maize Cell Genomics: Resources for Visualizing Promoter Activity and Protein Dynamics using Fluorescent Protein Lines*	10/14–9/18	66,480
National Science Foundation	*MaizeCODE—An Initial Analysis of Functional Elements in the Maize Genome*	6/16–5/19	183,073
National Science Foundation	*Biotechnology in American High Schools: Then and Now*	9/17–8/18	15,934
NONFEDERAL GRANTS			
Alfred P. Sloan Foundation	*DNA Center NYC* Start-up	12/13–6/19	62,457
Bank of America Charitable Foundation	Genetics and Biotechnology Lab	11/15–11/17	5,000
Beijing No. 166 High School	Chinese Collaboration Agreement	5/14–6/18	30,246
Breakthrough Prize Foundation	Laboratory Design and Teacher Training for Breakthrough Junior Challenge Prize Winners	12/15–12/17	187,232
Ashley and Frank O'Keefe	Support for Eastwood School and Greenvale School	12/16–12/17	5,000
Pinkerton Foundation	*Urban Barcode Research Program*	1/13–5/18	116,645
William Townsend Porter Foundation	*Harlem DNA Lab for Underprivileged Students*	4/17–3/18	13,500

The following schools and school districts each contributed $1,000 or more for participation in the *Curriculum Study* program:

Bellmore–Merrick Central High School District	$2,000	Long Beach Union Free School District	$3,000
East Meadow Union Free School District	$3,000	Massapequa Union Free School District	$3,000
East Williston Union Free School District	$1,500	North Shore Central School District	$2,000
Elwood Union Free School District	$2,000	Oceanside Union Free School District	$2,000
Fordham Preparatory School	$2,000	Oyster Bay–East Norwich Central School District	$2,000
Half Hollow Schools Central School District	$2,000	Plainedge Union Free School District	$2,000
Harborfields Central School District	$2,000	Plainview–Old Bethpage Central School District	$2,000
Herricks Union Free School District	$2,000	Portledge School	$3,000
Huntington Union Free School District	$2,000	Port Washington Union Free School District	$2,000
Island Trees Union Free School District	$2,000	Ramaz Upper School	$2,000
Jericho Union Free School District	$2,000	Roslyn Union Free School District	$2,000
Levittown Union Free School District	$2,000	Syosset Central School District	$3,000
Locust Valley Central School District	$2,000	Yeshiva University High School for Girls	$2,000

The following schools and school districts each contributed $1,000 or more for participation in the *Genetics as a Model for Whole Learning* program:

Adelphi STEP	$1,540	Jericho Union Free School District	$7,232
Bellmore Union Free School District	$10,420	Kings Park Union Free School District	$3,630
Bethpage Union Free School District	$5,235	Lawrence Union Free School District	$5,390
Cold Spring Harbor Central School District	$15,840	Locust Valley Central School District	$11,072
Commack Union Free School District	$2,530	Merrick Union Free School District	$7,055
East Williston Union Free School District	$2,117	Mott Hall II, New York	$1,155
Edgemont Union Free School District	$3,462	North Bellmore Union Free School District	$4,250
Elwood Union Free School District	$3,960	Northport–East Northport Union Free School District	$1,100
Evangelical Lutheran Church of the Redeemer, Glendale	$1,440	Oceanside Union Free School District	$1,800
Floral Park–Bellerose Union Free School District	$8,580	Oyster Bay–East Norwich Central School District	$2,640
Garden City Union Free School District	$10,872	Plainedge Union Free School District	$1,155
Great Neck Union Free School District	$9,075	Port Washington Union Free School District	$10,560
The Green Vale School	$1,622	PS 144, Forest Hills	$5,280
Greenwich Country Day School	$9,240	Rockville Centre Union Free School District	$10,824

†Includes direct and indirect costs.

Half Hollow Hills Central School District	$1,040
Hicksville Union Free School District	$1,540
Hofstra STEP	$1,650
Holy Child Academy	$1,705
Horace Mann School	$2,640
Huntington Union Free School District	$2,880
Island Park Public Schools	$2,160
Roslyn School District	$5,775
Scarsdale Public Schools	$6,924
Smithtown Union Free School District	$8,350
St. Patrick's School, Huntington	$2,160
Syosset Union Free School District	$44,385
Three Village Central School District	$3,080
Trevor Day School	$1,155

BANBURY CENTER GRANTS

Grantor	*Program*	*2017 Funding*
FEDERAL SUPPORT		
U.S. National Science Foundation, an award to R. Michelmore, University of California, Davis	Opportunities for Reduction of Aflatoxin Contamination of Food	$14,500
NIH/National Cancer Institute (grant to D. Tuveson, CSHL)	Integrated Translational Science Center Workshop	38,460
NONFEDERAL SUPPORT		
2Blades Foundation	NLRs Sans Frontières	5,000
Andrea B. and Peter D. Klein	Posttraumatic Neuroinflammation: Roles in Pathogenesis of Long-Term Consequences and Repair	40,000
Bill & Melinda Gates Foundation	Maximizing Impact of New HIV Prevention Technologies	60,111
Boehringer Ingelheim Fonds	BIF Fellows Retreat: Communicating Science	69,390
Burroughs Wellcome Fund	Ferroptosis: A Critical Review	15,000
Burroughs Wellcome Fund	NLRs Sans Frontières	5,000
CSHL Corporate Sponsor Program	Enhanceropathies: Enhancer function variation in animal development, morphological variation, and disease	48,335
CSHL Corporate Sponsor Program	Ferroptosis: A Critical Review	17,977
CSHL Corporate Sponsor Program	Neuropharmacology and Human Stem Cell Models	38,969
CSHL Corporate Sponsor Program	NLRs Sans Frontières	38,266
CSHL Corporate Sponsor Program	Opportunities for Reduction of Aflatoxin Contamination of Food	16,442
CSHL Corporate Sponsor Program	Regulated Necrosis, Pathways, and Mechanisms	18,945
Collaborative Medicinal Development, LLC	Ferroptosis: A Critical Review	15,000
DuPont Pioneer	NLRs Sans Frontières	5,000
Genentech	NLRs Sans Frontières	3,000
Genentech	Regulated Necrosis, Pathways, and Mechanisms	30,000
Gordon and Betty Moore Foundation	NLRs Sans Frontières	10,000
IC MedTech	Metformin: Translating Biology into the Clinic	25,000
Kansas State University	Opportunities for Reduction of Aflatoxin Contamination of Food	6,535
Keystone for Incubating Innovation in Life Sciences Network	Foundation2017: Bio-Entrepreneurship in NYC	35,500
L'Orèal USA	Chemiexcitation in Human Disease and Aging	4,535
Lieber Institute for Brain Development	Neuropharmacology and Human Stem Cell Models	13,070
Lustgarten Foundation	Lustgarten Scientific Advisory Board meeting	20,215
MARS, Inc.	Opportunities for Reduction of Aflatoxin Contamination of Food	15,000
Memorial Sloan Kettering Cancer Center	Ferroptosis: A Critical Review	2,550
Northwell Health	Project Santa Fe	35,265
Oliver Grace Chair Fund	Better Cancer Therapy from Redox Biology	74,797
Oliver Grace Chair Fund	Metformin: Translating Biology into the Clinic	40,000
Ono Pharmaceutical Co., Ltd.	Ferroptosis: A Critical Review	5,000
Steven and Alexandra Cohen Foundation	Protective Immunity and Vaccines for Lyme Disease	54,713
The Greater New York Chapter of The ALS Association	Cell Biology of ALS: Emerging Themes from Human Genetics	51,898
The LEO Foundation of Ballerup, Denmark	Chemiexcitation in Human Disease and Aging	41,797

CORPORATE SPONSOR PROGRAM FOR MEETINGS SUPPORT

Contributions from the following companies provide core support for the Cold Spring Harbor meetings program: Corporate Benefactor: Regeneron; Corporate Sponsors: Agilent Technologies, Bristol-Myers Squibb Company, Calico Labs, Celgene, Genentech, Inc., Thermo Fisher Scientific, Merck, Monsanto Company, New England Biolabs, Pfizer; Corporate Partners: Alexandria Real Estate, Gilead Sciences, Novartis Institutes for BioMedical Research, Sanofi.

The Laboratory acknowledges the generosity of the following companies that loaned equipment and reagents to the various courses: 3i; 10× Genomics; AB SCIEX LLC; ADInstruments Inc.; Advanced Analytical; Advanced Illumination; Agilent Technologies, Inc.; ALA Scientific Instruments; A-M Systems; Andor Technologies, Inc.; Applied Scientific Instrumentation, Inc.; BD Biosciences; Berthold Technologies USA, LLC; BioLegend; Bioline USA; Bio-Rad Laboratories; BioTek; Biotool; Bitplane; BMG Labtech; Bruker Daltonics, Inc.; Cairn Research Ltd.; Charles River Laboratories, Inc.; Chroma Technology Corporation; CNRS; Coherent; ConOptics; Covaris, Inc.; CrystaLaser; Diagenode; Drummond Scientific Company; eBioscience; EMD Millipore Corp.; Electron Microscopy Sciences; Epicentre Technologies Corp.; Eppendorf North America; Excelitas Technologies Corp.; Exiqon, Inc., Fotodyne Inc.; GE Healthcare; Hamamatsu Photonic Systems; Hamilton Drayage, Inc.; Hampton Research; Harlan; Harvard Apparatus, Inc.; Heka Instruments Inc.; Holoeye; Homebrew; Illumina, Inc.; Intan Technologies; Intelligent Imaging, Inc.; Invitrogen; Keyence; Kinetic Systems Inc.; Labcyte; Leica Biosystems; Leica Microsystems, Inc.; LI-COR; Life Technologies; Lumen Dynamics; Lumencor; MDS Analytical Technologies; MiTeGen; Molecular Devices Corp.; Morrell Instrument Co.; Nanodrop Technologies; NanoTemper Technologies; Narishige International USA, Inc.; Nasco; New England Biolabs, Inc.; Newport Corporation; Nikon Inc.; NSK America Corporation; OKO-Lab; PerkinElmer Life and Analytical Sciences; Photometrics; Physik Instruments; Precisionary Instruments; Prior Scientific; Promega Corporation; Qiagen Inc.; Quantitative Imaging Corp.; RC Testing Services; Sage Science, Inc.; Scientifica Ltd.; Singer Instruments; Sutter Instruments; Taconic Farms, Inc.; Tecan Group US, Inc.; The Jackson Laboratories; Thermo Fisher Scientific; Thor Labs; Tokai Hit; TotalLab Ltd.; Waters Corporation; White Labs; World Precision Instruments; Zenith Biotech.

DEVELOPMENT

With help from our loyal supporters, 2017 was an exceptional year for Cold Spring Harbor Laboratory (CSHL). For the first time, $7 million in unrestricted funding was raised from our annual appeal, golf tournament, Women's Partnership for Science lunch, and Double Helix Medals dinner. These vitally important funds provide flexibility to CSHL's President Bruce Stillman to invest in the most innovative research projects, which might not yet be eligible for federal funding.

Over the past several years, we have seen a tremendous growth in our Helix Society—a group of individuals who have named CSHL in their estate plans. Cynthia Stebbins (pictured below with fellow Helix Society member, Rod Cushman), an Honorary CSHLA Director and President's Council member, is a wonderful example of this.

As a neighbor and supporter of the Laboratory, I am a huge admirer of those young scientists who devote their lives to improving human health and well-being.

Decades of involvement as a member of the Laboratory Association Board of Directors and the President's Council have given me a real appreciation of just how complicated and difficult their work is.

Because the eventual impact of basic research is truly invaluable, I feel privileged to be able to establish a legacy that will help ensure their work and the future of Cold Spring Harbor Laboratory.

—Cynthia Stebbins

Most excitingly, renovation began this year on the historic 1953 Demerec Laboratory building. With a $25 million grant from New York State, it will be transformed into the state-of-the-art Center for Therapeutics Research. The $75 million total initiative will apply CSHL's biomedical expertise to advance therapeutics for genetic diseases. Governor Andrew Cuomo said, "New York is a leader in next-generation technology and sciences, and with the ground breaking of the new Center for Therapeutics Research in Cold Spring Harbor, we are supporting developments in research and medicine that will save lives."

This is an important time for science and especially for Cold Spring Harbor Laboratory. Thank you for partnering with us as we continue to transform human health.

Charles V. Prizzi, *Vice President for Development and Community Relations*

Cold Spring Harbor Laboratory Corporate Advisory Board

The Corporate Advisory Board (CAB) is composed of prominent business leaders from the tristate community and is a vital source of funding and outreach for Cold Spring Harbor Laboratory. Board members are the driving force behind the Laboratory's annual golf outing at Piping Rock Club, which raises critical unrestricted funding for research and education programs. CAB president Eddie Chernoff chaired the 24th annual CSHL outing, which honored CAB member and President of Harvest Real Estate and Timber Ridge Homes, Mark Hamer. The CAB members also participate in other events and fund-raisers for the Lab and are instrumental "ambassadors" to the community.

Corporate Advisory Board 2017
Edward A. Chernoff, MARS, *Chair*

Michael Aboff, Aboff's Inc.
Paul Amoruso, Oxford & Simpson Realty
Todd Andrews, Centerbrook Architects and Planners
Rocco S. Barrese, Dilworth & Barrese
Edward J. Blaskey, Sterling National Bank
Thomas J. Calabrese, Daniel Gale Sotheby's International Realty
Christopher Callaghan
John D. Catalano, Catalano Enterprises, LLC
Richard A. Catalano, KPMG, LLP
Jonathan Connors, Wells Fargo
Marian Conway, Ph.D., New York Community Bank Foundation
Philip D'Avanzo, Cushman & Wakefield
J. Daniel Decker, Eppendorf North America
Gregory DeRosa, Roanoke Holdings
Robert Dickstein, Ph.D.
Brian Fox, McKinsey & Company
Lawrence Goodman, Curtis, Mallet-Prevost, Colt & Mosle LLP
Mark W. Hamer, Harvest Real Estate Services, Inc.
Mary Jane Helenek
Richard W. Humann, H2M architects + engineers
Robert Isaksen, Bank of America
Alan L. Jakimo, Sidley Austin LLP
Patricia Janco-Tupper, Wells Fargo
John C. Kean III, Kean Development Company
Norman Kelker, Enzo Life Sciences
Amy Koreen, M.D.
Laurie J. Landeau, V.M.D.
Brian D. Lee, Newmark Grubb Knight Frank
David Lessing, Lessing's
Kyle Markland, Bethpage Federal Credit Union
Jeffrey L. Martin, M.D., Sight MD
Michael Maturo, RXR Realty LLC
Victoria Sagona Meagher
Stephen F. Melore, Farrell Fritz, P.C.
Frank T. O'Keefe, Infinitree LLC
Robert Palatnick, DTTC
John G. Passarelli, M.D.
David Peikon
Patricia Petersen, Daniel Gale Sotheby's International Realty
Joseph Roberto, BankUnited, N.A.
Gareth Roberts, HSBC
Matthew Roberts, NBTY
Stephen Ross, Ph.D., Nikon Instruments
Don Saladino, Drive 495
Raju Sarwal, M.D.
John Topolovec, TD Bank
Craig A. Weiss, T. Weiss Realty Corp.
Andrew Whiteley
Joseph M. Willen, Advantage Title

Golf Honoree Mark Hamer and his family with CSHL President Bruce Stillman and COO Dill Ayres

Cold Spring Harbor Laboratory Association

Under the leadership of the Cold Spring Harbor Laboratory Association (CSHLA) President, Michele Celestino, the CSHLA community raised $7 million dollars in vital, unrestricted funding to support the globally recognized research and education programs at CSHL. The CSHLA Board of Directors continues its efforts as ambassadors to the community at large. The Directors continue to host dinner parties in their homes for the Dorcas Cummings Symposium and participate in events. At this year's Women's Partnership for Science luncheon, we honored CSHLA Honorary Director, Freddie Staller. This sold-out event raised $200,000. The annual golf outing at Piping Rock was once again chaired by CSHLA Director Eddie Chernoff, and Mark Hamer was the honoree. This event raised more than $200,000. At the Double Helix Medals dinner at the American Museum of Natural History we honored long-time friends of the Lab Charles and Helen Dolan and journalist Tom Brokaw for their efforts in raising cancer awareness. This event raised $4.2 million.

In 2017, we also saw the successful conclusion of the 125th anniversary brick campaign that raised more than $180,000. The bricks will be installed in the newly renovated Beckman courtyard.

2017 Directors

Michele Munn Celestino, *President*
James A. Woods, *Treasurer*
Virginia Knott, *Secretary*
Elizabeth Ainslie
Lori Bahnik
Hans E.R. Bosch
Barbara W. Callaghan
Eric Carlstrom
Edward A. Chernoff
Maria H. de Lesseps
Nelson DeMille
Anna Saladino Gevinski
Gina Gherlone
David Goldring
Ronald A. Gottlieb
Nancy Lippman Israeli, M.D.
Joel J. Kassimir, M.D.
Kassahun Kebede
Peter J. Klein
Mickie J. Nagel
Eileen Otto
Whitney Posillico
Lonnie Shoff
Hope Geier Smith
Debbie Parmet Sondock
David Stark
Mary Striano
Nancy M. Szigethy
Peter S. Tilles
Lara S. Trafelet
Marjorie J. van de Stouwe, M.D.
Heather Warren Whitman

Honorary Directors

Mary D. Lindsay
Anne R. Meier
Cathy Cyphers Soref
Pearl F. Staller
Cynthia R. Stebbins

2017 CSHLA Directors

Honor Roll of Donors

Lifetime Contributions of $5 million+

Estate of Donald Everett Axinn
Dr. and Mrs. Lalit Rai Bahl
The Arnold and Mabel Beckman Foundation
Jamie Nicholls and Fran Biondi
David Boies, Chairman of Boies, Schiller & Flexner LLP
The Breast Cancer Research Foundation
Chan Zuckerberg Initiative DAF, an advised fund of Silicon Valley Community Foundation
Lavinia and Landon* Clay
The Dana Foundation
Mrs. Norris W. Darrell/Estate of Norris W. Darrell
Estate of Kathryn W. Davis
DeMatteis Family Foundation
Mr. and Mrs. Charles F. Dolan, Dolan Family Foundation
Martha Farish Gerry*, The William Stamps Farish Fund
Jacob Goldfield
Kate Medina Guthart and Leo A. Guthart
Howard Hughes Medical Institute
W.M. Keck Foundation
Mr. and Mrs. David H. Koch
Laurie J. Landeau Foundation
Mr. and Mrs. Robert D. Lindsay and Family
The Lustgarten Foundation
Nancy Abeles Marks
The G. Harold & Leila Y. Mathers Charitable Foundation
The Don Monti Memorial Research Foundation
New York State Empire State Development Corporation
Pershing Square Foundation
Cynthia Hazen Polsky and Leon Polsky, Hazen Polsky Foundation
The Quick Family
Charles and Marie Robertson Family
Drs. Marilyn and James Simons
The Simons Foundation
Alfred P. Sloan Foundation
St. Giles Foundation
The Stanley Medical Research Institute
The Starr Foundation
The Swartz Foundation
Thompson Family Foundation
Dr. and Mrs. James D. Watson
Mr. and Mrs. Henry Wendt III
Roy J. Zuckerberg Family Foundation

Contributions of $100,000+

Mr. and Mrs. Lee S. Ainslie III
Chen and Shao Anti-Cancer Foundation
Mr. and Mrs. Edward A. Chernoff, MARS
Davenport Family Foundation
Oliver S. and Jennie R. Donaldson Charitable Trust
Charitable Lead Annuity Trust under the Will of Louis Feil
Claire Friedlander Family Foundation
Genentech, Inc.
Lola Goldring
Michael J. Griffin
The Leona M. and Harry B. Helmsley Charitable Trust
Mr. and Mrs. Jeffrey E. Kelter
F.M. Kirby Foundation
Betsy and Bryan H. Lawrence
Louis Morin Charitable Trust
Dr. Robert W. Lourie
Mercer Family Foundation
William C. and Joyce C. O'Neil Charitable Trust
Estate of Thomas Parissidi
Amy and John Phelan
Alison Holtzschue and Douglas Schloss
Dr. and Mrs. James M. Stone
Mr. and Mrs. Paul J. Taubman
The Charles B. Wang International Foundation
Dr. and Mrs. Stuart T. Weisbrod

Contributions of $30,000+

The Bahnik Foundation
Mr. and Mrs. Hans E.R. Bosch
Broad Hollow Bioscience Park, Inc.
Stephen B. Burke, NBCUniversal
Cedar Hill Foundation
Children's Heart Foundation
Ellen and Casey Cogut
Mr. and Mrs. Roderick H. Cushman
Drs. Pamela Hurst-Della Pietra and Stephen Della Pietra
Dr. Lester Dubnick
Ike, Molly and Steven Elias Foundation
Fera Pharmaceuticals, LLC
Joni Gladowsky Breast Cancer Foundation
The Go Go Foundation
Dr. and Mrs. Philip Goelet, Francis Goelet Charitable Trust
Mr. and Mrs. Ronald A. Gottlieb
The Marc Haas Foundation
Mr. and Mrs. Mark W. Hamer
The Irving A. Hansen Memorial Foundation
David and Ginny Knott
Tomislav Kundic
Mary D. Lindsay
The Meier and Linnartz Family Foundation
Breast Cancer Awareness Day in memory of Elizabeth McFarland

Governor Andrew Cuomo (*fourth from right*), New York State Senator Carl Marcellino (*second from right*), and other New York state representatives joined CSHL President Bruce Stillman (far *left*) for the ground breaking of the Center for Therapeutics Research. New York state granted the Laboratory $25 million toward the total $75 million project, which includes renovation of the historic Demerec Laboratory.

*Deceased

(*Left to right*) CSHL Chairman Jamie Nicholls, Honorary Trustee Jim Simons, and Trustee George Yancopoulos at the Double Helix Medals Dinner.

Dr. and Mrs. Howard L. Morgan
National Basketball Association
Mr. and Mrs. Michael Pascucci, Sebonack Golf Club
Dr. and Mrs. John G. Passarelli, SightMD
Michelle Paternoster Foundation
Pfizer, Inc.
Thomas C. Quick
Regeneron Pharmaceuticals
Christina Renna Foundation
Charles and Marie Robertson Foundation
Thomas M. Rutledge, Charter Communications
Joshua Sapan, AMC Networks
Susan Wasserstein and George Sard
Mr. and Mrs. Thomas A. Saunders III
Edith and Alan Seligson
Mr. and Mrs. Theodore Smith III, The Geier Foundation
Mr. and Mrs. Howard Solomon
Mr. and Mrs. Erwin P. Staller
Swim Across America Nassau/Suffolk
Friends of TJ Foundation
The Tow Foundation
The Wasily Family Foundation

Contributions of $10,000+

Paul G. Allen Family Foundation
Alnylam Pharmaceuticals, Inc.
Mr. and Mrs. W. Dillaway Ayres, Jr.
Deborah C. Benjamin
Mr. and Mrs. Vincent A. Breitenbach
The Broder Family Foundation
Dr. David P. Brown
Mr. and Mrs. Christopher C. Callaghan
Mr. and Mrs. Eric Carlstrom
Mr. and Mrs. James J. Celestino
Dr. and Mrs. Bayard Clarkson
Katya H. Davey
Michel David-Weill
Mr. and Mrs. Geoffrey de Lesseps, Alpha Omega Charitable Foundation
Barbara J. Amonson and Vincent J. Della Pietra
Sandy and Nelson DeMille
Laura and John Desmarais
Drs. Miki and David Donoho
Fortunato Sons Contracting, Inc.
Mr. and Mrs. Robert A. Gay
Elbridge T. Gerry, Jr.
Mr. and Mrs. Albert Gherlone
The Gilder Foundation
H2M architects + engineers
The Randall and Mary Hack Foundation
Susan T. Harris
Hearts Who Care/Hanging with Heroes
Yaoqing & Shouying Huang Family Foundation
Joan Morthland Hutchins
Elizabeth McCaul and Frank Ingrassia
Drs. Nancy and Ron Israeli
Annette Kade Charitable Trust
Kappa Alpha Order's Breastfest
Dr. and Mrs. Joel J. Kassimir
Andrea B. and Peter D. Klein
Eugenia McGrath and Thomas Korossy
Gordon Lamb, The Jefferson Family Charitable Foundation
Mr. and Mrs. Stephen M. Lessing
Dr. and Mrs. Arthur D. Levinson
Mr. and Mrs. Thomas H. Lister
Mr. and Mrs. John Macaskill, Macaskill Foundation
Edward E. Matthews
Dr. Marcia Kramer Mayer
Mr. and Mrs. Edward C. Meagher, Carter, DeLuca, Farrell & Schmidt LLP
Gillian and Eduardo Mestre
The Tom and Judy Moore Foundation
The Morningside Foundation
Mr. and Mrs. Douglas P. Morris
Northwell Health
University of Notre Dame
Susanne Olin
Wendy O'Neill
Mr. and Mrs. Jonathan Otto
David P. Pearson
Patricia J. Petersen, Daniel Gale Sotheby's International Realty
Estate of Margaret Schierberl Pomeroy
William Townsend Porter Foundation

(*Left to right*) Judy Cormier, CSHL President Bruce Stillman, Chairman Jamie Nicholls, and Tom and Amanda Lister at the Double Helix Medals Dinner.

*Deceased

CSHL Trustee Robert Lourie (*fifth from left*) and guests at the Double Helix Medals Dinner.

Mrs. Richardson Pratt, Jr. and family, The Spionkop Charitable Trust
Thomas and Marina Purcell Family Foundation
Drs. Marjorie van de Stouwe and Scott Ratner
Rauch Foundation
Mr. and Mrs. John R. Reese
Ride For Life, Inc.
Pamela and Richard Rubinstein Foundation
Mr. and Mrs. Herbert J. Siegel
Steven J. Simmons
Mr. and Mrs. Harry Slatkin
Kristin and Dustin Smith
Deborah K. Solbert
Mr. and Mrs. Douglas S. Soref
David Stark, Teva Pharmaceutical Industries
Mr. and Mrs. James F. Stebbins
Dr. and Mrs. Bruce W. Stillman
The Lauri Strauss Leukemia Foundation

Golf honoree and Corporate Advisory Board member Mark Hamer (*left*) with CSHL President Bruce Stillman (*right*) at the Golf Tournament.

Mr. and Mrs. Vincent T. Striano, Striano Electric
Sullivan & Cromwell, LLP
Nancy Szigethy
Anne and John Tatta Family Foundation
Anne D. Thomas
Mr. and Mrs. Peter S. Tilles
Joan P. Tilney
Mr. and Mrs. Edward Travaglianti
Mr. and Mrs. Paul A. Vermylen, Jr.
West Islip Breast Cancer Coalition for L.I.
Ann Eden Woodward Foundation
Mr. and Mrs. Robert C. Wright
The Bradley Zankel Foundation
The Zilkha Foundation

Contributions of $5,000+

Dinu Albeanu
Fred and Janet Baron
Michael and Victoria Bershadsky
Mr. and Mrs. Timothy S. Broadbent
Mr. and Mrs. Thomas J. Calabrese, Jr.
Mr. and Mrs. George W. Carmany III
Centerbrook Architects and Planners, LLC
Mr. and Mrs. Robert W. Chasanoff
Rita M. Cleary
Dr. Marian Conway, New York Community Bank Foundation
Mr. and Mrs. James C. Cook
Denise R. Coyle
Dau Family Foundation
Philip D'Avanzo, Cushman & Wakefield
Mr. and Mrs. Claudio Del Vecchio
Lee MacCormick Edwards Charitable Foundation
Mr. and Mrs. Mark Eichorn
Diane and Bob Fagiola
J.C. Flowers Foundation
Michael L. Focazio, Carissa Maringo Fund
Putney Cloos and Brian Fox
Carolyn and Leonard Gero
Glen Cove C.A.R.E.S.
Pierre and Paula Gonthier Family Foundation
Google
Robin Hadley
The Hastings Foundation
Jo-Ellen and Ira Hazan
Mr. and Mrs. Alan L. Jakimo
Patricia Janco-Tupper and Rebecca Creavin, Wells Fargo Private Bank
John C. Kean, Kean Development Company, Inc.
Kassahun Kebede
Errol S. Kitt, GEI Consultants, Inc.
The Klinsky Family Charitable Fund
KPMG LLP
Lazer, Aptheker, Rosella & Yedid, P.C.
Mr. and Mrs. Richard Leeds
David Lessing, Lessing's Hospitality
Robert T. Lindsay
Smadar Maduel-Chaluts
Mrs. David J. Mahoney
Robert L. Marcus
Kyle Markland, Bethpage Federal Credit Union
Masthead Cove Yacht Club
Mr. and Mrs. Michael Maturo
Irene McPhail, Cablevision Alumni/ Leadership Alliance International
Steven Melore, Farrell Fritz P.C.
Mr. and Mrs. Jeffrey A. Nagel
Nature's Answer Foundation
Newmark Grubb Knight Frank
NorthMarq Capital
Mr. and Mrs. Frank T. O'Keefe
Mr. and Mrs. Robert Palatnick
Seemon H. and Natalie E. Pines Foundation
Platsky Company, Inc.
Mrs. Joseph D. Posillico, Jr.
Dr. Amy Rappaport Koreen and Dr. Judith Rappaport
Tom and Marianne Reifenheiser
Joe Roberto, BankUnited, N.A.
Lari and Frank Russo
Lonnie and Michael Shoff
Owen and Bernadette Smith
John Sobolewski
Mr. and Mrs. Josh Spoerri
Lee Steinberg
Three Strohm Sisters Family Foundation
Mr. and Mrs. F. Bradford Townley
Mr. and Mrs. Stanley S. Trotman, Jr.
United HealthCare Services, Inc.
Vector Vitale
John B. Vermylen
Craig A. Weiss
Mr. and Mrs. W. Fifield Whitman

(*Left to right*) Jim Watson, John Reese, and Michael O'Brien at the Helix Society lunch held in June. The Helix Society is made up of individuals who have named Cold Spring Harbor Laboratory in their estate plans.

Joseph M. Willen, Advantage Title Agency, Inc.
Mr. and Mrs. Steven Jay Winick

Contributions of $1,000+

Dustin Aaronson
Nancy and Michael Aboff
American Society of Plumbing Engineers
Paul F. Amoruso, Oxford & Simpson Realty Services, Inc.
Mr. and Mrs. Paul Annunziato
Mr. and Mrs. William R. Araskog
Debra and Scott Arenare
Paul and Mary Auersperg
AvalonBay Communities
Hal and Tricia Avidano
Mr. and Mrs. Henry D. Babcock
Deanna Bahel
Mrs. Robert H.B. Baldwin
Bank of America
Baranay Family Foundation
Elizabeth M. Beach
Mr. and Mrs. Dominic P. Bencivenga
Kathy Walsh and Gene Bernstein
Elizabeth Bertani
Don and Jan Bettencourt
Frank J. Biondi, Jr.
Drs. Renee A. Fitts and David Botstein
Dr. William Braden III, The Porter Braden Fund of CCF
Katharine Bradford Foundation
Dr. Steven Brenner
Brookhaven National Laboratory
Mr. and Mrs. Matthew J. Bruderman
Josephine Marrali and Richard Bryant
Mr. and Mrs. W. Russell Byers, Jr.
The Bynum Family Foundation
C.M.S. Electrical Services, Inc.
John Callaghan, Callaghan Lucerino & Associates LLP
Ann and Peter Cannell
Canon Solutions America, Inc.
Carol and Stephen Canter
CardWorks Servicing, LLC
Martha M. Carlstrom
Louise M. Parent and John Casaly
Mr. and Mrs. John D. Catalano
Mary and Richard Catalano
Susan and Glenn Cohen
Mr. and Mrs. Jonathan Connors
Ruth Cooper
Oliver Cope
Core Group Architects, LLP
Cosel-Pieper Family Foundation
Cunniff Family Fund
Gregory DeRosa
Dilworth & Barrese, LLP
Mr. and Mrs. Todd A. DiScala
Karen L. Dissinger
Mr. and Mrs. Jack Dolce
Mr. and Mrs. Joseph T. Donohue
Thomas J. Dowling, Dominick & Rose Ciampa Foundation
DRK Real Estate Consulting
Electronix Systems Central Station Alarms, Inc.
Frederick W. Engel
Enzo Life Sciences, Inc.
Fred and Barbara Fisher
The Flinn Family Charitable Fund
Adam Friedlander
Dr. and Mrs. Candido E. Fuentes-Felix
Drs. Sydney Gary and Alexander Gann
E. Maxwell Geddes, Jr.
Nina and Carl Girolamo
Drs. Joan and Alfred Goldberg
Mr. and Mrs. William H. Grover
Thomas C. Gsell
Paul Hamer
Robert S. Hamer
Mr. and Mrs. Robert S. Hamer
Hamlet Golf and Country Club and ClubCorp Charities
Reverend Linda Peyton Hancock
Margaret M. Hargraves
Ed Harlow
The Hartman Family Foundation
Mary Jane Helenek
William Allan Herzog
James B. Hicks, Ph.D.
Robert and Cindy Higginson
Theodora W. Hooton
Frederic B. Ingraham
Walter B. James Fund No. 2
Johnston Family Foundation
Mr. and Mrs. Barclay G. Jones III
Kleinknecht Family Foundation
Mr. and Mrs. Ragnar Meyer Knutsen
KWM CPAS LLP
Dr. David Lane
Mr. and Mrs. James M. Large, Jr.
Henry and Marsha Laufer
Levin, Shea & Pfeffer
The Litwin Foundation, Inc.
Long Island 2 Day Walk to Fight Breast Cancer
Kevin G. Mahony
Heather and Robert Maione
Mallinson Family Fund
Mr. and Mrs. James W. McCormack
Rusty and Peggy McCormack
Allison Maddox and Charles McCurdy
Michael McGlone
Mr. and Mrs. Neil T. McGoldrick
Mr. and Mrs. Ronald P. McNavich
The Elena Melius Foundation
Mr. and Mrs. Clarence F. Michalis
Micro Strategies Technology Solutions
Mr. and Mrs. Ray Mohr
Walter and Sheila Mondschein
Dr. Kathleen Mooney
Emil and Irene Moshkovich
Mr. and Mrs. Stephen V. Murphy
Richard P. and Faye Nespola Charitable Foundation
Norwood Foundation, Inc.
Patricia and Hugh O'Kane, Jr.
Mr. and Mrs. Richard S. Overton
Tracy Dellomo and Clifford Packingham
Rhoda Parmet
Jeffery K. Parsigian
Mr. and Mrs. Nicholas B. Paumgarten
Mr. and Mrs. David Peikon
Peto Maintenance and Development, LLC
Mr. and Mrs. Simon Pharr
Dr. and Mrs. Anthony F. Piazza
Pilkington Family Fund
Christina and Grant Porter
Praxis Public Relations, Inc.
Nicole and Charlie Prizzi
Mr. and Mrs. John Pufahl
Mr. and Mrs. Thomas L. Pulling
Barbara and Preston Pumphrey
Racanelli Construction Company, Inc.
Katie Raftery
Mr. and Mrs. Harold Rechler
Bernard and Anne Reynolds
RGM Insurance Agency, Inc.
Gareth Roberts, HSBC Bank USA

CSHL President Bruce Stillman (*center*) and Senior Fellow Sarah Diermeier (*left*) with Honorary CSHLA Director Freddie Staller (*right*) at the Women's Partnership for Science lunch. Freddie was honored at this year's event, and Sarah spoke about her breast cancer research at the Laboratory.

Mr. and Mrs. William D. Roche
Arthur M. Rogers, Jr.
Richard Rose
Dr. and Mrs. Gerald M. Rubin
The Ruder Family Foundation
RXR Co Property Management
DiDi and Neal Ryland
Dr. and Mrs. Raju Sarwal
Mr. and Mrs. Peter G. Schiff
Robin and Enrique Senior
Robert D. Sires
Debbie Parmet Sondock
Mr. and Mrs. James L. Spingarn
The Spring Family Fund
Stand Up For Suzanne
Syosset Limousine Service, Inc.
Judith and Irwin Tantleff
Mrs. David S. Taylor
TD Bank, N.A.
Professor Jeremy W. Thorner
Pamela M. Thye
Mr. and Mrs. Pascal Franchot Tone
Trinity College
Doreen Mochrie and Christopher Tuohy
UBS Financial Services
Barbara Ulman
Dr. and Mrs. Paul Ulmer
Uni-Data & Communications, Inc.
Universe Kogaku (America), Inc.
Diane Valente
Richard John Walk
Mr. and Mrs. Thomas J. Walsh
Mr. and Mrs. Douglas A. Warner III
Deborah Norville and Karl Wellner
Westerman Ball Ederer Miller Zucker & Sharfstein, LLP
Gerald I. White
Dr. and Mrs. Michael Wigler
Erby and Marcia Wilkinson
Sheldon Gerry Withers
Women's Financial Group, Inc.

In-Kind Gifts

Annabel Romero
Mr. and Mrs. W. Dillaway Ayres, Jr.
Banfi Vintners
Bank of America
Pien Bosch
Brooks Brothers
Cold Spring Harbor Wine Shoppe
Estée Lauder
Fortunoff Fine Jewelry
Cathy Freudenberg Art
Garagetek
Mr. and Mrs. Albert Gherlone
Girl Obsessed
Harlan Estate
J. McLaughlin
Jonathon's Ristorante
Kimberly Gorman Muto Photography
Living
Nest Fragrances
Dr. and Mrs. John Passarelli
Penguin Random House
Red
Safavieh Home Furnishings
Sandbar
Raju Sarwal, M.D.
Tom Schaudel
Sedoni Gallery
Dr. Bruce Stillman
Eleanor Stone NYC
Jan Tozzo
Glass.Pieces.Studio
Marjorie van de Stouwe, M.D.
Tres Jolie Salon
Mr. and Mrs. Howard J. Tytel
Wempe
Windham House

Contributions in honor of

The Baccellieri Family
Jamie Nicholls and Fran Biondi
Eddie and Marc Chernoff
Susan Cohen
Tony Coles
Susan and Jim Cronley
Kristina Perkin Davison
Sue Dillof
Ron Gottlieb
Mark Hamer
Stephen Lerner
Mary D. Lindsay
Alea Mills
Mind Over Moon
Barry P. Newburger
Patricia Quick
Serena Reichel
Justin Rosati
Richard Slawsky
Freddie Staller
Bruce Stillman
Jill Stone
Sallie and Alex van Rensselaer
Michael Wigler
Evelyn Witkin
Jan Witkowski

Contributions in memory of

Sandy Aranoff
Morton and Becky Aronson
TJ Arcati
Carl Bobick
Jack and Mildred Bowers
Ellen Brenner
Carol A. Buonaiuto
Vincent Carey
John P. Cleary
Daniel W. Coakley
Estelle Cole
Norris Darrell
Maryliz Dickerson
Tina Duque
Claire Martin Fairman
Bud Friedman
Nancy Slaughter Gay
Joni Gladowsky
Christopher G. Goff
Isabelle Grenier
Michael Guttenberg
Teresa Haire
Henry Jost
Gilbert Kahn
Donald Kuhn
Heidi La Spina
Gabriella Lasasso
Betty Long
Susan Petersen Lubow
William Manos
Carissa Maringo
Frank McGirr
John J. McGowan
Linda Mitgang
Goldie Moradi
Frank Morelli
Gerald Olin
Marilyn Palatnick
Lawrence Parker
Edward C. Pitre II
Edith Platsky
Yvette Rafiy
Marion Reagan
Norman Remler
Elizabeth Roberts
John Roosevelt
Rona Ross
Eric Rudowski
Diane Edmin Sachs
Rolene Schliff
Miriam Schlofsky
Andrew Seligson
Suzanne Shea
Marilyn Sitrin
Mallory Smith
Jason Stewart
Mrs. Sylvia Taubman
Melvin and Shirley Thorner
Laurie Ostroff White
Daniel Wolf
Mark Zoller

LABORATORY MANAGEMENT

Bruce W. Stillman, Ph.D., President & CEO
W. Dillaway Ayres, Jr., Chief Operating Officer

RESEARCH

David L. Spector, Ph.D.
Director of Research

Walter L. Goldschmidts, Ph.D.
Vice President, Executive Director of Sponsored Programs

Research Operations

Sydney Gary, Ph.D.
Director, Research Operations

Professors

Douglas Fearon, M.D.
Hiroyasu Furukawa, Ph.D.
Thomas Gingeras, Ph.D.
Professor and Head Functional Genomics
Z. Josh Huang, Ph.D.
David Jackson, Ph.D.
Leemor Joshua-Tor*, Ph.D.
Adam Kepecs, Ph.D.
Alexei Koulakov, Ph.D.
Adrian Krainer, Ph.D.
Program Chair, Cancer and Molecular Biology
Bo Li, Ph.D.
Zachary Lippman, Ph.D.
Robert Maki, M.D., Ph.D.
Robert Martienssen*, Ph.D.
Program Chair, Genomics and Plant Biology
W. Richard McCombie, Ph.D.
Alea Mills, Ph.D.
Partha Mitra, Ph.D.
Adam Siepel, Ph.D.
Program Chair, Simons Center Quantitative Biology
Marja Timmermans, Ph.D.
Nicholas Tonks, Ph.D.
Deputy Director, CSHL Cancer Center
Lloyd Trotman, Ph.D.
David Tuveson, M.D., Ph.D.
Director, CSHL Cancer Center
Linda Van Aelst, Ph.D.
Michael Wigler, Ph.D.
Anthony Zador, M.D., Ph.D.
Program Chair, Neuroscience

Associate Professors

Dinu Albeanu, Ph.D.
Gurinder Atwal, Ph.D.
Anne Churchland, Ph.D.
Mikala Egeblad, Ph.D.
Jesse Gillis, Ph.D.
Christopher Hammell, Ph.D.
Ivan Iossifov, Ph.D.
Alexander Krasnitz, Ph.D.
Pavel Osten, M.D., Ph.D.
Stephen Shea, Ph.D.
Raffaella Sordella, Ph.D.
Arne Stenlund, Ph.D.
Christopher Vakoc, M.D., Ph.D.
Doreen Ware**, Ph.D.

Assistant Professors

Camila dos Santos, Ph.D.
Tatiana Engel, Ph.D.
Molly Hammell, Ph.D.
Justin Kinney, Ph.D.
Je Hyuk Lee, M.D., Ph.D.
Dan Levy, Ph.D.
Gholson Lyon, M.D., Ph.D.
David McCandlish, Ph.D.
Vincent Pedmale, Ph.D.
Jessica Tollkuhn, Ph.D.
Hongwu Zheng, Ph.D.

Research Professor

Darryl Pappin, Ph.D.
Scott Powers, Ph.D.

Research Assistant Professor

Kenneth Chang, Ph.D.
Scott Lyons, Ph.D.
Jonathan Preall, Ph.D.
Michael Ronemus, Ph.D.
Johannes Yeh, DDS, Ph.D.
Seungtai Yoon, Ph.D.

Senior Computer Scientist

Boris Yamrom, Ph.D.

Computer Scientist

Osama El Demerdash, Ph.D.
Vivek Kumar, Ph.D.

Senior Research Investigator

Zihua Wang, Ph.D.

Research Investigators

Joan Alexander, M.D.
Gayatri Arun, Ph.D.
Christopher Faehnle, Ph.D.
Qing Gao, Ph.D.
Carla Gauss, Ph.D.
Sara Goodwin, Ph.D.
Gilbert Henry Jr., Ph.D.
Manzar Hossain, Ph.D.
Yoon-Ha Lee, Ph.D.
Youngkyu Park, Ph.D.
Michael Regulski, Ph.D.
Yi-Jun Sheu, Ph.D.
Damianos Skopelitis, Ph.D.
Mona Spector, Ph.D.
Priya Sridevi, Ph.D.
Asya Stepansky, Ph.D.
Herve Tiriac, Ph.D.
Kai Yu, Ph.D.
Huiqing Zhan, Ph.D.

Manager, Clinical & Translational Research

Jessica Gotterer

Professor Emeritus

James D. Watson, Ph.D.
Oliver R. Grace Professor Emeritus/Chancellor Emeritus

Adjunct Professors

Grigori Enikolopov, Ph.D.
Gregory Hannon, Ph.D.
James Hicks, Ph.D.
Scott Lowe, Ph.D.
Larry Norton, M.D.

Adjunct Associate Professors

Michael Schatz, Ph.D.
Glenn Turner, Ph.D.

Visiting Scientists/ Collaborative Scientists

Ebru Demir, Ph.D.
Hyun Jae Pi, Ph.D.

Visiting Professors

Matthew Ridley, Ph.D.
Stuart Weisbrod, Ph.D.

QB Visiting Scholar

Bhubaneswar Mishra, Ph.D.

Director, Gene Targeting Shared Resources

Leyi Li, Ph.D.

Senior Fellow

Dannielle Engle, Ph.D.

Clinical Fellow

Kenneth Yu, M.D.

CSH Fellow

Semir Beyaz, Ph.D.
Jason Sheltzer, Ph.D.
Lingbo Zhang, Ph.D.

Translational Fellow

Fieke Froeling, M.D., Ph.D.

Postdoctoral Fellows

Sandra Ahrens, Ph.D.
Jean Albrengues, Ph.D.
Xu An, Ph.D.
Lindsey Baker, Ph.D.
Balasooriya Balasooriya, Ph.D.
Sara Ballouz, Ph.D.
Leah Banks, Ph.D.
Prasamit Baruah, Ph.D.
Walter Bast, Ph.D.
Pascal Belleau, Ph.D.
Yael Berstein, Ph.D.
Kuhulika Bhalla, Ph.D.
Sonam Bhatia, Ph.D.
Sonali Bhattacharjee, Ph.D.

*Employee of the Howard Hughes Medical Institute
**Employee of the United States Department of Agriculture

Giulia Biffi, Ph.D.
Amit Blumberg, Ph.D.
Christopher Bonham, Ph.D.
Marion Bosc, Ph.D.
Christian Bravo-Rivera, Ph.D.
Robert Bronstein, Ph.D.
Federico Carnevale, Ph.D.
Irene Casanova Salas, Ph.D.
Hong Goo Chae, Ph.D.
Chen Chen, Ph.D.
Cristina Chen, Ph.D.
Wei Chia Chen, Ph.D.
Xiaoyin Chen, Ph.D.
Yushu Chen, Ph.D.
Quentin Chevy, Ph.D.
Tsung Han Chou, Ph.D.
Tae Mun Chung, Ph.D.
Hannes Claeys, Ph.D.
Megan Crow, Ph.D.
Samantha Cyrill, Ph.D.
Chand Parvez Danka Mohammed, Ph.D.
Juliane Dassler Plenker, Ph.D.
Filipe de Sousa Borges*, Ph.D.
Edgar Demesa Arevalo, Ph.D.
Hanfei Deng, Ph.D.
Roman Dvorkin, Ph.D.
Elad Elkayam*, Ph.D.
Ela Elyada, Ph.D.
Mary Feigman, Ph.D.
Matthew Fisher, Ph.D.
Daniel Fuerth, Ph.D.
Akihiro Funamizu, Ph.D.
Alessandro Furlan, Ph.D.
Debarati Ghosh, Ph.D.
Nicholas Gladman, Ph.D.
Luana Guerriero, Ph.D.
Priyanka Gupta, Ph.D.
Rasmani Hazra, Ph.D.
Xueyan He, Ph.D.
Anat Hendelman, Ph.D.
Longwen Huang, Ph.D.
Yifei Huang, Ph.D.
Yuhan Huang, Ph.D.
Dhananjay Huilgol, Ph.D.
Stellar Hur, Ph.D.
Change-Il Hwang, Ph.D.
Jonathan Ipsaro, Ph.D.
Anbalagan Jaganathan, Ph.D.
Matt Jaremko, Ph.D.
Yinping Jiao, Ph.D.
Ashley Juavinett, Ph.D.
James Jung*, Ph.D.
Matthew Kaufman, Ph.D.
Virginie Kergourlay, Ph.D.
Gukhan Kim, Ph.D.
Munenori Kitagawa, Ph.D.
Olaf Klingbeil, Ph.D.
Hideki Kondo, Ph.D.
Choon-Tak Kwon, Ph.D.
Seung Cho Lee, Ph.D.
Young Koung Lee, Ph.D.
Zachary Lemmon, Ph.D.
Bo Li, Ph.D.
Jiayun Li, Ph.D.
Qian Li, Ph.D.
Shujing Li, Ph.D.
Siran Li, Ph.D.
Jayon Lihm, Ph.D.
Kuan-Ting Lin, Ph.D.
Bodu Liu, Ph.D.
Lei Liu, Ph.D.
Bin Lu, Ph.D.
Jiangteng Lu, Ph.D.
Wai Kit Ma, Ph.D.
Grinu Mathew, Ph.D.
Katherine Matho, Ph.D.
Luis Mejia Jr., Ph.D.
Olivia Mendivil Ramos, Ph.D.
Koji Miyabayashi, M.D., Ph.D.
Andrea Moffitt, Ph.D.
Hemanth Mohan, Ph.D.
Rodrigo Munoz Castaneda, Ph.D.
Adriana Munoz Jimenez, Ph.D.
Simon Musall, Ph.D.
Farzaneh Najafi, Ph.D.
Arun Narasimhan, Ph.D.
Farhana Nasrin, Ph.D.
Gal Nechooshtan, Ph.D.
Ellen Louise Noren Lindback, Ph.D.
Dawid Nowak, Ph.D.
Kin On*, Ph.D.
Torben Ott, Ph.D.
Jean-Sebastien Parent, Ph.D.
Sapan Patel, Ph.D.
Dennis Plenker, Ph.D.
Mohammad Rahman, Ph.D.
Michael Regan, Ph.D.
Ashlan Reid, Ph.D.
Jie Ren, Ph.D.
Benjamin Roche, Ph.D.
Daniel Rodriguez-Leal, Ph.D.
Jae Seok Roe, Ph.D.
Nikolay Rozhkov, Ph.D.
Elena Rozhkova, Ph.D.
Thiago Santos Gouvea, Ph.D.
Christine Scaduto, Ph.D.
Juergen Scharner, Ph.D.
Andrea Schorn, Ph.D.
Mario Shields, Ph.D.
Atsushi Shimada, Ph.D.
Timothy Somerville, Ph.D.
Sebastian Soyk, Ph.D.
Sarah Starosta, Ph.D.
James Sturgill, Ph.D.
Lijuan Sun, Ph.D.
Xueqin Sun, Ph.D.
Johanna Syrjanen, Ph.D.
Yilin Tai, Ph.D.
Nami Tajima, Ph.D.
Julian Taranda, Ph.D.
Yusuke Tarumoto, Ph.D.
Claudia Tonelli, Ph.D.
Michael Tramantano, Ph.D.
Raditya Utama, Ph.D.
Anand Vasudevan, Ph.D.
Prabhadevi Venkataramani, Ph.D.
Bo Wang, Ph.D.
Bor-Shuen Wang, Ph.D.
Minghui Wang, Ph.D.
Xingang Wang, Ph.D.
Zhikai Wang, Ph.D.
Yiliang Wei, Ph.D.
John Wilson, Ph.D.
Sze Wong, Ph.D.
Qingyu Wu, Ph.D.
Xiong Xiao, Ph.D.
Fang Xu, Ph.D.
Xiaosa Xu, Ph.D.
Tao Yang, Ph.D.
Zhaolin Yang, Ph.D.
Min Yao, Ph.D.
Dinar Yunusov, Ph.D.
Xian Zhang, Ph.D.
Juannan Zhou, Ph.D.
Imanol Zubiete Franco, Ph.D.

Graduate Students

Lyndsey Aguirre
Brinda Alagesan
Naishitha Anaparthy
Batuhan Baserdem
Giorgia Battistoni
Benjamin Berube
Brianna Bibel
Anna Lien Bouhuis
Sara Boyle
Emilis Bruzas
Jason Carter
Kung-Chi Chang
Lital Chartarifsky
Derek Cheng
Hsiang-Chen Chou
Alberto Corona
Talitha Forcier
William Galbavy
Nicholas Gallo
Jordan Gegenhuber
Sanchari Ghosh
Hamza Giaffar
Jacqueline Giovanniello
Kristina Grigaityte
Bradford Gulko, Ph.D.
Michael Gutbrod
Abram Handly-Santana, Ph.D.
Benjamin Harris
Kelly Hills-Muckey
Yixin Hu
Yuzhao Hu
Elizabeth Hutton
Dong-Woo Hwang, Ph.D.
Shruti Iyer
David Johnson
Daniel Kepple
Young Jin Kim
Navyateja Korimerla
Christopher Krasniak
Matthew Lee
Jesse Levine
Shaina Lu
Devon Lukow
Dennis Maharjan
Diogo Maia e Silva, M.D.
Laura Maiorino
Paul Masset
Katarina Meze
Judith Mizrachi
Manojit Mosur Swamynathan
Alexandra Nowlan
Kathryn O'Neill
Tobiloba Oni
Tzvia Pinkhasov
Sashank Pisupati
Sofya Polyanskaya
Zhe Qian
Ricardo Raudales
Suelynn Ren
Deborah Rupert
Danilo Segovia
Luqun Shen
Padmina Shrestha
Sergey Shuvaev
Dongyan Song
Olya Spassibojko
Martyna Sroka
Colin Stoneking
Jennifer Thalappillil
Ngoc Tran
Jue Xiang Wang
Xiaoli Wu
Cole Wunderlich
Wenbo Xu
Yali Xu
Ran Yan
Georgi Yordanov
Allen Yu
Chengxiang Yuan
Anqi Zhang

*Employee of the Howard Hughes Medical Institute

Qian Zhang
Siwei Zhang, Ph.D.

Computational Science

Peter Andrews, Ph.D.
Mihail Bota, Ph.D.
Kapeel Chougule
Astrid Deschenes
Alexander Dobin, Ph.D.
Evan Ernst
Daniel Ferrante, Ph.D.
Elena Ghiban
Yuan Hao, Ph.D.
Aashna Jhaveri
Ying Jin, Ph.D.
Jude Kendall
Melissa Kramer
Sunita Kumari, Ph.D.
Xu Li
Zhenyuan Lu
Steven Marks
Matthew Moss
Demitri Muna, Ph.D.
Andrew Olson
Ritika Ramani
Manthan Shah
Joshua Stein, Ph.D.
Oliver Tam, Ph.D.
Ammar Tareen, Ph.D.
Marcela Tello-Ruiz, Ph.D.
James Thomason III
Kannan Umadevi Venkataraju
Liya Wang, Ph.D.
Xiaofei Wang, Ph.D.
Xuehong Wei
Matthew Wroten, Ph.D.
Fengyuanshan Xu
Miu ki Yip

Research Associates

Carmelita Bautista
Barry Burbach
Hilary Cox
Jorg Drenkow
Hyun Soo Kim, Ph.D.
Lisa Manche
Saikat Nandi, Ph.D.
Ramesh Palaniswamy
Anirban Paul, Ph.D.
Linda Rodgers
Suzanne Russo
Caizhi Wu, M.D.
Melody Wu, Ph.D.
Scilla Wu
Lifang Zhang, Ph.D.

Scientific Project Managers

Kathleen Kelly, Ph.D.
Alexander Vaughan, Ph.D.

Laboratory Facility Managers

Sabrina Boettcher
Amy Brady
Patricia Brady
Julie Cheong
Robert Eifert
Nour El-Amine
Stephanie Goldsmith*
Calvin Lee
Pamela Moody
Timothy Mulligan
Keith Rivera
Joseph Simorowski*
Spencer Teplin
Dennis Thomas, Ph.D.
Ming Wang

Scientific IT Support

David Trimboli
Peter Van Buren

Scientific Administrators

Lorraine Baldwin
AnnMarie Ballew
Christine Bedell
Patricia Bird
Jessica Bisceglio
Barbara Cascone
Melissa Daley
Becky Dong
Susan Fredericks
Irene Gill
Denise Hallinan
Lisa Kimoto
Idee Mallardi
Lindsay Monaco
Sandra Moran
Kylie Parker
Dorothy Tsang Hu
Madeline Wisnewski

Research Technicians

Valeria Aguillon Rodriguez
Alexandra Ambrico
Ariana Amendolara
Mirko Andreoli, Ph.D.
Andalus Ayaz
Sandra Barrera
Brian Benz
Eleni Boikoglou, Ph.D.
Denise Cahn
Yon Chang
Joshua Cohen
Devon Cowan
Jonathan Crain
Jenesis Curtis
Justin Dalrymple
Cassidy Danyko
Martin Davis
Michael Dworjan
Patrick Einhorn
Amanda Epstein*
Margaret Falkowski
Christy Felice
Carolina Maria Fernandes Henriques
Connor Fitzpatrick
John Gardiner IV
Steven Georgiou
Elizabeth Gibson
Steven Gluf
Inessa Hakker
Kaarina Hanington
Fangle Hu
Qingting Hu, M.D.
Eitan Joshua
Cassandra Kanzler
Clancy Kelahan
Anup Khanal
Maja Kornaj
Andrew Krainer
Eunjung Lee
Yan Li
Ying Hsiu Liu
Landon Lodato
Edward Lu
Kathleen Lucere
Beicong Ma
Michelle Ma
Arthur Makarenko
Senem Mavruk Eskipehlivan
Joseph Milazzo
Kristin Milicich
Stephanie Muller
David Ng
Aigoul Nourjanova
Joseph O'Rourke
Jie Ouyang
Ann Palladino
Hardik Patel
Paola Pisterzi
Umamaheswari Ramu
Dylan Rebolini
Claire Regan
Michael Riggs
Julie Rosenbaum
Polona Safaric Tepes
Stephen Savoia
Christina Schoepfer
Charles Shaw
Regina Shaw
Noriko Simorowski
Tara Skopelitis
Benjamin Spielman
Logan Stecher
Yu-Chi Sun
Alexander Tolpygo
Claire Uhlman
Maria Cristina Velasquez Cobos
Kamalakannan Vishnu
Jing Wang
Robert Wappel
Kaitlin Watrud
Simone Weinmann
Shih Ting Yang
Melissa Yao
Catherine Ye

Laboratory Facility Staff

Rachel Strittmatter, DVM
Director of Lab Animal Resources/Attending Veterinarian
Alexander Aguilera
Damir Begonja
Lisa Bianco
William Blatz
Erin Brosnan
Oral Bryan
Amy Burrows
Michael Cahn
Manuel Canas
Jodi Coblentz
Steven Collins
Kristin Cox
Charlene De Poto
Maria Demott
Jared Downing
Rhonda Drewes
Eileen Earl
Wilson Garciaguirre
Susan Guterman
Jill Habel
Tristan Jarrett
Alyssa Molloy
Maria Mosquera
Gustavo Munoz
Jason Pellegrini
Mason Salese
Ann Smith
Laura Stonebridge
Walter Vigil Garcia
Peter Wimett Jr.

*Employee of the Howard Hughes Medical Institute

Media Makers/Farm Assistants

Nancy Bolanos
Terrance Chisum
Adriana Hincapie
Kelly Ann Holfester
Vincent Meschan
Michele Murrell
Shouyun Qiao
Amy Qiu
Sarah Vermylen

EDUCATION

Banbury Center

Rebecca Leshan, Ph.D.
Director
Michelle Corbeaux
Patricia Iannotti
Barbara Polakowski

CSHL Press

John R. Inglis, Ph.D.
Executive Director
Barbara Acosta
Janet Argentine
Kevin John Black
Cynthia Blaut
Carol Brown
Kathleen Bubbeo
Kathleen Cirone
Laureen Connell, Ph.D.
Marie Cotter
Jennifer DeLeon, Ph.D.
Corrisa Farmer
Miriam Fein, Ph.D.
Bibiane Garite
Michael Henigman
Nancy Hodson
Deborah Jarski
Catherine Kelly
Tara Kulesa
Dana Macciola
Wayne Manos
Mala Mazzullo
Joanne McFadden
Mary Mulligan
Stephanie Novara
Stephen Nussbaum
Dorothy Oddo
Robert Redmond
Theodore Roeder
Mary Sabala
Eric Sawey, Ph.D.
Diane Schubach
Richard Sever, Ph.D.
Inez Sialiano
Marcie Siconolfi
Maria Smit, Ph.D.
Hillary Sussman, Ph.D.
Linda Sussman
Jesse Ulloa
Denise Weiss

DNA Learning Center

David A. Micklos
Executive Director
Lindsay Barone, Ph.D.
Elna Carrasco
Lauren Correri
Heather Cosel-Pieper
Alison Cucco
Jenna Dorey
Cristina Fernandez Marco, Ph.D.
Cornel Ghiban
Mary Lamont
Susan Lauter
Melissa Lee
Christine Marizzi, Ph.D.
Amanda McBrien
Pauline McGlone
Colleen McKechnie
Valerie Meszaros
Eric Nash, Ph.D.
Michael Paul
Sharon Pepenella, Ph.D.
Jorge Perez de Acha Chavez
Keil Thomas
Jason Williams
Chun-hua Yang
Xiaoqun Zhang

Library and Archives

Ludmila T. Pollock
Executive Director
Paula Abisognio
Thomas Adams
Tara Bonet-Black
Clare Clark
Matthew Covey, Ph.D.
Matthew Dunn, Ph.D.
Elizabeth Pessala, J.D.
Stephanie Satalino
Gail Sherman

Meetings & Courses Program

David J. Stewart, Ph.D.
Executive Director
Stephen Alexander
Alfred Baman
Margaret Bernhardt
Gerard Burlandi
Edward Campodonico
Cathleen Carr
Joseph Carrieri
Meredith Cassuto
Betty Cavallaro
Reginald Cayo
Katiria Ceballos
Maoyen Chi, Ph.D.
Wendy Crowley
Matthew Denninger
William Dickerson
Jane Donaldson, Ph.D.
Alicia Franco
Bradley Frey
Constance Hallaran
Patricia Klieger
Charla Lambert, Ph.D.
Rachel Lopez
Dennis McGuire
Andrew Mendelsohn
Nicholas Moore
Maureen Morrow
Andrea Newell
Kenneth Orff
Valerie Pakaluk
Jonathan Parsons
Victor Ryder Sr.
Stephen Silver
Mary Smith
Erica Stang
Samuel Stewart
Vivian Vuong
James Whitaker
Barbara Zane

Office of the President

Terri I. Grodzicker, Ph.D.
Dean of Academic Affairs/ Editor, Genes & Development
Jan A. Witkowski, Ph.D.
WSBS Professor

Watson School of Biological Sciences

Alexander Gann, Ph.D.
Lita Annenberg Hazen *Dean*
Kimberly Creteur
Kimberley Geer
Alyson Kass-Eisler, Ph.D.
Monn Monn Myat, Ph.D.

ADMINISTRATIVE DEPARTMENTS

Finance and Accounting

Lari C. Russo
Chief Financial Officer
Sharon Bense
Elise Best
Karine Boyce
Joseph Conti
Damian Desiderio
Maryanne Detweiler
Kristin Dunckley
Karen Filasky
Mary Ellen Goldstein
Evelyn Imbriani
Mary Mastropaolo
Alison McDermott
Denise Miller
Kristine Murphy
Patricia Penner
Nicole Solomon
Patricia Urena
Gisela Webb
Ann Williams

Central Administration

Margie Bell
Maureen Berejka
Carol Gundersen
Bridgette Johnson
Carol Link
Denise Roberts, Ph.D.

Culinary Services

James Hope
Director
Doris Abrego
Inolka Abreu
Ana Adames
Alan Aguirre
Jeannette Amato
Eliseo Amaya
Ingrid Amaya
Yamic Ambra
Julie Barkan
Maeve Barth Dwyer
Francis Bowdren
Maryam Brinkwart
Elmer Canales
William Carmona
Dilicia Castro Najera
Francis Cayamcela
James Charvis
Marvin Chavez
Rene Chavez Contreras
Christina DeDora
Erica Douglas
Julie Ehrlich
Suzanne Esposito
Bartola Garcia
Luis Garcia
Andres Gonzalez
Silvia Gonzalez

Jennifer Gordon
Kamal Johnson
Deborah Koslosky
Pamela Lancellotti
Jennifer LoSordo
Juan Magana
Jose Maradiaga Gonzalez
Thomas McIlvaine
Laurie Moller
Jason Moss
Frederik Munter
Jose Parada Alvarado
Joel Pena Corvera
Jose Reyes
Jesus Rivera
Susana Roman
Marlenis Romero
Robert Scherer
Karen Scola
Shannon Silva
Jose Suarez
Jose Velasquez
Hector Vigil Garcia
Maria Villatoro
Maria Vilorio
Tracy Woolfson
Briana Yim
Jennifer Ziegler
Lidia Zuleta

Development

Charles V. Prizzi
Vice President, Development/ Community Relations
Diane Fagiola
Bridgette Johnson
Karen Orzel
Jill Stone

Facilities

Arthur Brings
Vice President, Chief Facilities Officer
Robert Albinson
Leslie Allen
Peter Anderson
Lucas Arevalo
Jerry Armstrong
William Bishop
Ralph Byers
Caitlin Carroll
Dessie Carter
Daniel Chapman
Anthony Ciccarella
Christopher Clark
Robert Collins
Drew Comer
Peter Dale
John Damico
Nicholas Dassau
GianPaolo Dato
James Dean
Robert Dickerson
Glen DiMaria
Kevin Donohue
Sarah Drayzen
Wesley Dreusike
Paul Edwards
Elias Eifert
Rafael Estrella
Matthew Ferraro
Vito Gassi
Robert Gensel
Patrick Ginty
Erick Greene
Leo Greene
Stephen Gregorovic
Philip Grella Jr.
Wilman Gutierrez
Wayne Hamilton
Gerard Holler
Jason Holzman
Adam Hopper
Joseph Houser
Matthew Hoyer
Christopher Hubert
Louis Hunter
Michael Hutchinson
Randal Jones
Daniel Jusino
Kevin Keenan
Jeffrey Klaverweiden
Robert Kopetic Jr.
Mason Kraese
Paulo Krizanovski
Joseph Leonard
Thomas LeRoy
Walter Leute
Jian Lin
Charles Mandell
Gregory Marotta
Patricia McAdams
Tahid McCoy
Christopher McEvoy
Kevin McHugh
Daniel Miller
Susan Montalvo
Oscar Moreira
Guido Mosquera
Mary Muno
Michael Murphy
Kenneth Nastri
Herbert Neazer
Michael Nolasco
Wayne Pav
Christopher Pearson II
Jose Pena Corvera
Saul Pena Corvera
John Pettit Jr.
John Pisciotta
Manuel Ramones
Wilson Ramones
Steven Randall
Elisha Rendal Sestoso
Jaqueline Reynolds
Alfredo Rivera Jr.
Ronald Romani
Frank Russo
Richard Rutherford
Jeffrey Sanchez
Eric Sappell
Steven Sardone
John F. Schindler
John L. Schindler
Claudia Schmid
Lauren Schmidt
Peter Schuman
John Shea
Cesar Sisalima
John Sundman
Steven Tang
Hock Yew Teo
Leigh Tyrrel
James Van Wie
Fredy Vasquez
Trinidad Velasquez
Wilfredo Velasquez
Benjamin Veneable III
Kyle Welton
Leslie Wenzel
Tracey Anne White
Graham Wildt
Daniel Zimmer

Human Resources

Katherine G. Raftery
Vice President
Andres Alarcon
Wendy Alexander
Camille Cava-Belluscio
Patricia Crawford
Jessica Heckel
Lesley Inglis
Laura Magri
Laura Moran
Elsy Padro
Victoria Panebianco
Patricia Schoeffler
Samantha Walbridge
Lauren Warmbrand
Rita Wienclaw

Information Technology

Hans Erik Aronson
Vice President, Chief Information Officer
Alex Barton
Edward D'Azzo-Caisser
Darian Defalco
Daniel Drayzen
Frank Falco
Kelvin Flores
Todd Heywood, Ph.D.
Thomas Karow
Leila Kelly
Sean Kelly
Caryn Koza
Michael Malave
Louis Malfi
Adam Mills
Bhargav Patel
Barbara Peters
Vincent Reale
Salvatore Serafino
Michael Skuthan
David Jonathan Torres
Steven Zielinski

Procurement

Michael Marchesiello
Vice President
Deborah Aufiero
Kimberly Bronson
Susan De Angelo
Jeffrey DuPree
Elizabeth Janow
Anne Knoch
Christopher Oravitz
Barbara Purcell
Jorge Ramirez
Krystyna Rzonca

Public Affairs

Dagnia Zeidlickis
Vice President, Communications
Andrea Alfano
Margot Bennett
Jessica Giordano
Philip Renna
Sara Roncero-Menendez
Susan Runkowski
Brian Stallard
Peter Tarr, Ph.D.

Research Operations

Rachel Desrochers

Diane Esposito, Ph.D.
Research Compliance/Research Investigator
Robert Gerdes
Jaclyn Jansen, Ph.D.
Laura Lynn
Julie Sutherland

Sponsored Programs

Walter L. Goldschmidts, Ph.D.
Vice President, Executive Director of Sponsored Programs
Irina Armstrong
Nicole DeVito
Carol DuPree
Diane Errico
David Garner
Donna Gavin
Jill Hemish, Ph.D.
Joanne Janelli
Ruta Ledins
Hans Lindback
Josephine Marrali, J.D.
Dave Neale
Catherine Perdikoylis
Bailey Stark
Jaclyn Tenaglia

Technology Transfer

Teri Willey
Vice President, Business Development/Technology Transfer
Fauzia Chaudhary, Ph.D.
Elizabeth Cherian-Samuel
Vladimir Drozdoff, J.D., Ph.D.
Michael Gurtowski
Elizabeth Hand
Radhakrishnan Narayanan, Ph.D.
Elizabeth Panagot

Legal

Debra Arenare, J.D.
Vice President, General Counsel
Alexander Sosnovski

RESEARCH STAFF DEPARTURES DURING 2017

Adjunct Professor

Yi Zhong

Database Manager, Genomics

Christopher Zaleski

Senior Fellow

Sarah Diermeier

Computer Scientist

Ashwin Prakash

Senior Research Investigator

Carrie Davis

Clinical Research

Madison Miller

Research Investigator

Navasona Krishnan
Vasily Vagin

Visiting Scientists/ Collaborative Scientists

Ilaria Cascone
Chand Parvez Danka Mohammed
Delphine Fagegaltier
Nicholas Gladman
Marco Grillo
Maria Gutierrez-Nava
Yinping Jiao
Xiangning Li
Hysell Oviedo
Natalia Peunova
Justin Robert Prigge
Ileng Ramachandran
Jason Tucciarone
Christopher Winefield
Ling Zhang

Visiting Clinical

Richard Burkhart
Jose Vilaro

Postdoctoral Fellows

Michael Campbell
Michelle Carmell
Iok In Chio
Monika Chugh
Sarah Diermeier
Max Doerfel
Gaofeng Fan
Michael Feigin
Wesley Frey
Rowan Herridge
Aman Husbands
Byoung Il Je
Steffen Knauer
Victoria Kuettner
Bryan Lanning
Michael Ludwig
Mehrab Modi
Shinichi Nakamuta
Gonzalo Hugo Otazu Aldana
Katherine Petsch
Arnaud Pommier
Nalani Sachan
Hillary Schiff
Carlos Stahlhut Espinosa
Marcus Stephenson-Jones
Ante Tocilj
Cao Xu
Qiangqiang Zhang

Computational Science

Wen-Wei Liao

Research Associates

Diana Ravens
Patricia Wendel